VISIT…

LANZAROTE Caliente.com

目　录

第一部分　绪　论

第一章　研究的定义、过程及分类 …………………… (3)

第一节　“研究”的界定 …………………… (3)
第二节　研究过程 …………………… (6)
第三节　研究的分类 …………………… (9)
小结 …………………… (13)
讨论题 …………………… (14)

第二章　基本概念 …………………… (15)

第一节　假设、理论及理论模式 …………………… (15)
第二节　研究对象总体与样本 …………………… (19)
第三节　变　　量 …………………… (20)
第四节　测量方式 …………………… (25)
第五节　如何将理论定义变为操作定义? …………………… (29)
小结 …………………… (32)
讨论题 …………………… (33)

第二部分　研究方法

第三章　提出研究问题 …………………… (37)

第一节　提出问题的步骤 …………………… (37)
第二节　选题、定题过程中的困难 …………………… (43)
小结 …………………… (48)
讨论题 …………………… (48)

第四章　阅读文献 …………………… (50)

第一节　文献来源 …………………… (50)

第二节　查询、阅读、综述文献的一般程序 ……………… (53)
第三节　确定文献阅读范围 ……………………………… (56)
第四节　整理与综合文献 ………………………………… (58)
小结 ……………………………………………………… (60)
讨论题 …………………………………………………… (60)

第五章　选择研究设计 ………………………………… (62)

第一节　定量与定性 ……………………………………… (62)
第二节　研究问题和研究设计的关系 …………………… (67)
第三节　单一式研究设计还是混合式研究设计 ………… (69)
第四节　设计分类的复杂性 ……………………………… (73)
小结 ……………………………………………………… (75)
讨论题 …………………………………………………… (76)

第六章　问卷调查 ……………………………………… (77)

第一节　问卷调查简介 …………………………………… (77)
第二节　问卷设计 ………………………………………… (77)
第三节　问卷答案的分级 ………………………………… (88)
第四节　选择调查对象 …………………………………… (91)
第五节　实施问卷调查 …………………………………… (96)
小结 ……………………………………………………… (98)
讨论题 …………………………………………………… (98)

第七章　实验研究 ……………………………………… (100)

第一节　什么是实验研究? ……………………………… (100)
第二节　因果关系 ………………………………………… (102)
第三节　实验的效度 ……………………………………… (103)
第四节　实验研究的种类 ………………………………… (110)
第五节　实验研究的程序 ………………………………… (114)
小结 ……………………………………………………… (118)
讨论题 …………………………………………………… (119)

第八章　个案研究 ……………………………………… (120)

第一节　什么是个案研究 ………………………………… (120)

第二节　选择研究对象 ……………………………………（121）
第三节　收集数据 ………………………………………（123）
小结 ……………………………………………………（137）
讨论题 …………………………………………………（137）

第九章　微变化研究 ……………………………………（138）

第一节　什么是微变化研究？ ……………………………（138）
第二节　认知心理学领域中微变化研究的案例 …………（141）
第三节　微变化研究在二语习得领域内的应用前景 ……（145）
小结 ……………………………………………………（153）
讨论题 …………………………………………………（153）

第十章　统计学入门 ……………………………………（154）

第一节　统计学简介 ……………………………………（154）
第二节　描述性统计 ……………………………………（155）
第三节　推断性统计 ……………………………………（163）
小结 ……………………………………………………（167）
讨论题 …………………………………………………（167）

第十一章　定量数据分析 ………………………………（168）

第一节　SPSS for Windows 概述 ………………………（168）
第二节　问卷数据分析 …………………………………（170）
第三节　分析实验数据 …………………………………（192）
小结 ……………………………………………………（194）
讨论题 …………………………………………………（195）

第十二章　定性数据分析 ………………………………（196）

第一节　分析定性数据的 3 个选择 ……………………（196）
第二节　数据的准备 ……………………………………（198）
第三节　定性分析 ………………………………………（202）
第四节　效度和信度 ……………………………………（212）
小结 ……………………………………………………（212）
讨论题 …………………………………………………（213）

第三部分　论文写作

第十三章　概　述 ………………………………………… (217)

第一节　论文是 thesis 还是 dissertation? …………… (217)
第二节　论文结构 ……………………………………… (217)
小结 ………………………………………………………… (224)
讨论题 ……………………………………………………… (224)

第十四章　论文写作 …………………………………… (226)

第一节　"引言"的撰写 ……………………………… (226)
第二节　"文献回顾"的撰写 ………………………… (227)
第三节　"研究方法"的撰写 ………………………… (234)
第四节　"研究结果与讨论"的撰写 ………………… (238)
第五节　如何写"结论" ……………………………… (243)
小结 ………………………………………………………… (247)
讨论题 ……………………………………………………… (247)

第十五章　论文的写作格式 …………………………… (248)

第一节　论文写作格式的基本要求 ………………… (248)
第二节　美国心理学协会论文写作规则 …………… (249)
小结 ………………………………………………………… (260)
讨论题 ……………………………………………………… (261)

参考文献 ………………………………………………… (262)

附录

附录一：A SURVEY STUDY ……………………………… (271)
附录二：实验研究 ……………………………………… (290)
附录三：个案研究 ……………………………………… (299)
附录四：随机数字表 …………………………………… (317)

第一部分 绪 论

绪论包括两章。第一章介绍“研究”的定义、过程及分类，第二章解释相关的基本概念。通过绪论的学习，你应该能：

- 理解“研究”的基本标准；
- 了解“研究”全过程；
- 区别不同类型的“研究”；
- 掌握与“研究”相关的基本概念。

第一章 研究的定义、过程及分类

第一节 “研究”的界定

应用语言学研究者 E. Hatch & H. Farhady（1982）将“研究”定义为：“用系统方法探求问题的答案”（a systematic approach to finding answers to questions）。这个定义涵盖了研究的 3 个要素：问题（questions）、系统方法（a systematic approach）和答案（answers）。本书将采用 Hatch 和 Farhady 的定义，系统描述研究方法和指导论文写作。

图 1.1 根据上述定义描述了研究的简单模式。所有的研究都始于问题，解决问题的手段是系统的研究方法，最终得到的结果是问题的答案。有了适当的问题，并不能自动产生系统的研究方法，没有系统的方法自然不会得到有效的答案。退一步说，即使找到值得研究的问题，又采用了系统的方法，答案有时也未必有效，因为研究结果的解释可能会有悖逻辑，或经不起推敲。因此，从事研究时，3 个要素必须等量齐观，不能厚此薄彼。

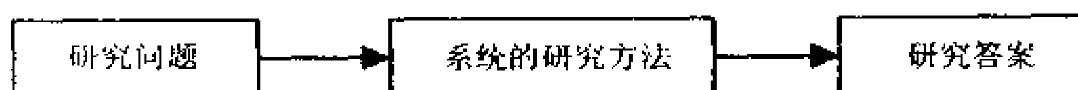


图 1.1：研究模式简图

1. 确定研究问题

确定研究问题是研究的第一步。然而要找到令人满意的问题并非易事。所谓令人满意的问题必须满足 3 个条件：有价值（significant）、有新意（original）、有可行性（feasible）。

所谓有价值，指的是实用价值或理论价值。前者指研究结果能对实

践活动的改进提出建议；后者指对现存理论提出挑战、支持或补充。一项研究可能具有实用价值或者理论价值，或者两者兼而有之。但不少缺乏研究经验的人时常提出价值不大的问题。例如，有人提出：“在二语写作中，写作水平高的人所犯语法错误是否少于写作水平低的人?”显然，写作水平高的人肯定比写作水平低的人要少犯语法错误，否则何以称作写作水平高的人呢？如果把问题改成“写作水平高的人与写作水平低的人相比，他们在语法准确性方面有何不同?”就有价值了。这样的研究结果一方面可以帮助教师了解写作水平高的人与写作水平低的人在语法准确性上存在的差异；另一方面这些结果可以为建立中介语发展的理论模式提供证据。

所谓问题有新意，是一个相对的概念，它存在于完全没有新意到很有新意的一个连续区间内。完全没有新意的题目可能有两种。一种是前人在这个领域中已经研究得非常透彻，第二种是所提的问题纯属常识性问题，答案不究自明。具有百分之百新意的问题是指没有文献可查考，没有方法可借鉴，一切都需要研究者自己去尝试的问题。笔者认为完全没有新意和具有百分之百新意的问题都不适合硕士生和博士生做研究。比较合适的新意程度是：提出的问题与前人的研究有些相同，又有些不同。不同点可以体现在一个或多个方面，例如学习者成绩上的差别有的是因为学习环境不同，有的是因为学习者的类型不同，有的是因为收集数据和分析数据的方法不同。

所谓有可行性，指的是研究者能在规定的时间范围内，利用现有的人力、物力、财力找到问题的答案。这一要求似乎有点多余，然而初搞研究的人往往对这一点认识不足，常常有好高骛远的毛病。

爱因斯坦曾经说过：“提出一个问题往往比解决一个问题更为重要。解决一个问题也许只需要数学上的运算或实验的操作，而提出新的问题，新的可能性，从新的角度去看旧的问题却需要创造性的想像力，这标志着科学的真正进步。”因此，确定一个好的问题是研究中最难的一部分，而这一点往往被人们忽视。研究就好比长途跋涉，问题确定得不好，就会误入歧途，结果既浪费时间又浪费精力。难怪 Nunan（1992：211）在书中强调“只要能把研究问题确定好，花多少时间都值得。”

2. 运用系统的研究方法

本书所阐述的应用语言学研究是社会科学领域内教育科学和语言科

学的交叉性学科研究。众所周知，语言现象、教育过程错综复杂，如果不能运用适当的科学方法，进行系统的研究，往往就无法得到有信度、有效度、有价值的发现。

系统的研究方法是一种科学的、公认的、可操作的、可重复的研究方法。它需要一整套明确、合理的研究步骤。这些步骤的合理性决定了研究结果的可信度。如果是采用定量设计，这些步骤必须在收集数据之前确定，并要通过先导性试验，对不完善的地方进行调整后，再进行正式研究，以确保设计步骤的可行性。如果是采用定性设计，具体步骤通常在研究过程中逐步确定。这两种情况，不管哪一种，选择研究对象的步骤、收集数据的过程和分析数据的方法都应该有理有据，清清楚楚记录在案。换句话说，研究过程中的所有步骤都是公开、透明，可表述的，每个步骤都应该有科学的依据和合理的解释，而不是研究者随心所欲的决定。

运用系统的研究方法从事研究也是社会科学中进行重复研究的必要条件。如果一项研究设计的步骤含混不清，别人就无法重复，无法验证，缺乏可比性。这一点在我国还没有得到足够的重视。自然科学实验结果的发表一般是多次试验的结果。而社会科学研究的对象是千差万别的人，我们不可能在同一批受试者身上多次重复某一项试验，因此社会科学的研究结果必须要由其他研究者进行重复检验，只有在多次重复研究的基础上，我们才能建立比较可信的理论。

3. 寻求有效度的答案

作为一项研究，就必然希望寻求一个有效度的答案。所谓问题的答案要有效度，就是说答案是惟一的答案。如果还有其他答案，这项研究的效度就要受到怀疑。这里我们不妨来看两个例子。例如：某项研究的问题是："阅读量与词汇量是否有密切的关系？"研究结果发现阅读量与词汇量之间存在着正相关，换句话说，二语学习者阅读量越大，他们的词汇量就可能越大。这一研究结果看起来既符合我们的直觉又合乎逻辑。然而，一位有经验的研究者很快发现该项研究中所用的词汇测试存在严重问题，这时答案的效度就值得怀疑，因为这个问题现在有了两个答案：（1）阅读量与词汇量之间的确存在着密切关系；（2）阅读量与词汇量可能没有密切关系，但由于词汇测试的偏差，产生了正相关的关系。再如，有一项研究旨在研究二语学习中的性别差异。研究发现在英

语水平测试中，英语专业的女生比男生考得好，研究者由此得出结论：在二语学习方面，女生比男生聪明。很明显，关于这一现象，我们可以有十余种解释，如：女生比男生花的时间多；女生比男生把考试当回事；最聪明的男生去学理工科，而没有学英语专业，等等。既然其他的解释也讲得通，那么，这一结论的效度就值得怀疑。

第二节 研究过程

人们常常从不同的角度，用不同的方法来观察研究过程。这里，我们介绍的“循环说”和“流程说”就是其中的两种。

1. 循环说

Rudestam & Newton（1992）认为研究过程的各个阶段是循环往复的。简单地说，研究过程至少由两个循环组成，每个循环有四个阶段（参见图 1.2）。第一个循环的各阶段依次为：进行实际观察、提出假设、建构理论框架、产生具体选题。第二个循环的各阶段依次为：收集数据、分析数据、把研究成果纳入某一理论框架，再确定新的研究选题。Rudestam 和 Newton 认为归纳和演绎应该贯穿研究的全过程，大致说来左边的各步骤使用的是归纳，右边的各步骤使用的是演绎。

第一循环

第一循环始于实际观察，通过观察，确定研究方向；第二阶段提出假设，即以陈述形式对变量之间的关系进行描述，如：学习目的与选择学习策略有关；第三阶段根据相关理论和前人的研究，提出一个理论框架，然后把假设置于理论框架内考察。初做研究的人觉得这一阶段很难、要求很高；第四阶段产生具体的研究选题，选题一旦确定，第一个循环就完成了。

第二循环

第二循环的第一阶段是针对要研究的问题收集数据，这一过程也可以看作为实际观察的一种形式。第二阶段是分析数据。第三阶段形成理论框架，即在分析数据的过程中，研究者通过归纳进行概括，然后把概括出的东西形成理论框架。第四阶段根据理论框架，提出新的研究问题，并进一步明确研究的意义。

"循环说"强调研究的动态性和循环性。但这个模型以定量研究为基础，没有考虑定性研究的实际情况。定量研究是为了检验假设，验证理论；而定性研究初期不可能形成完整的假设。我们以后再对这个问题进行讨论。

2. 流程说

Wen（2001）提出的"流程说"（参见图1.3）比较简单，易于理解，更适合研究新手。它强调的是研究全过程应该完成的任务和步骤，而不是这些任务或步骤之间的相互作用。根据图1.3，研究包括6习任务：确定选题、阅读文献、选择研究设计、收集数据、分析数据、撰写论文。这里用向下的箭头表示6大任务之间的顺序。最后的返回箭头从"撰写论文"又回到"确定选题"，表示论文完成后，研究者将会投入新的研究。

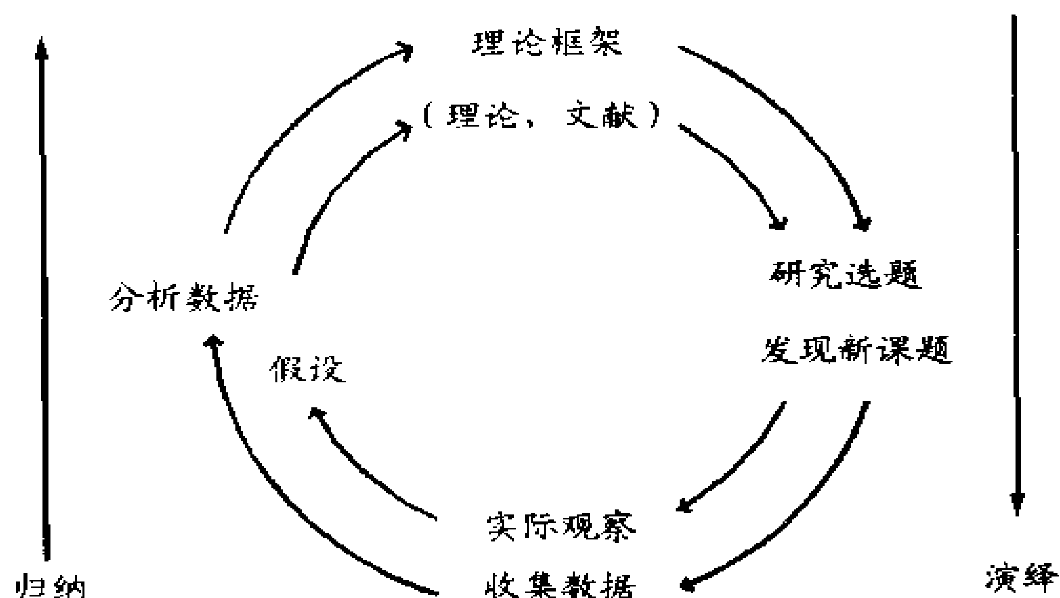


图 1.2：循环说模型

在研究过程中，这6项任务不能互相替代，不能厚此薄彼，它们是一个有机的整体。其中任何一项任务完成得不好，都有可能毁掉整个研究。这6项任务之间存在着一定的逻辑关系。一般说来，前后顺序不能改变。例如：收集数据不能在确定选题前面，又如：撰写论文不能在分

析数据前面。但在实际操作过程中，各项任务之间的顺序不一定总是清楚明了的。例如阅读文献贯穿于选题和选择设计研究的始终。当研究者对研究方向有个大致了解后，就需要通过查找阅读相关文献来缩小研究范围、确定研究焦点。选择研究设计时，还需要阅读文献，这是因为选择研究设计不仅受到选题的制约，而且还受到以前研究方法的限制。

作为研究新手，最好先按照这个顺序逐一完成任务，一旦熟悉了研究的过程，就可以灵活处理，不受这个顺序的限制。Blaxer（1996）曾经指出："现实中的研究往往杂乱无章。"因此从这个意义上说，流程图没有真实反映研究现实。但流程图有两个好处：一是简明扼要，能够帮助研究入门者记住研究的大致过程，二是既适合定量研究，又适合定性研究。

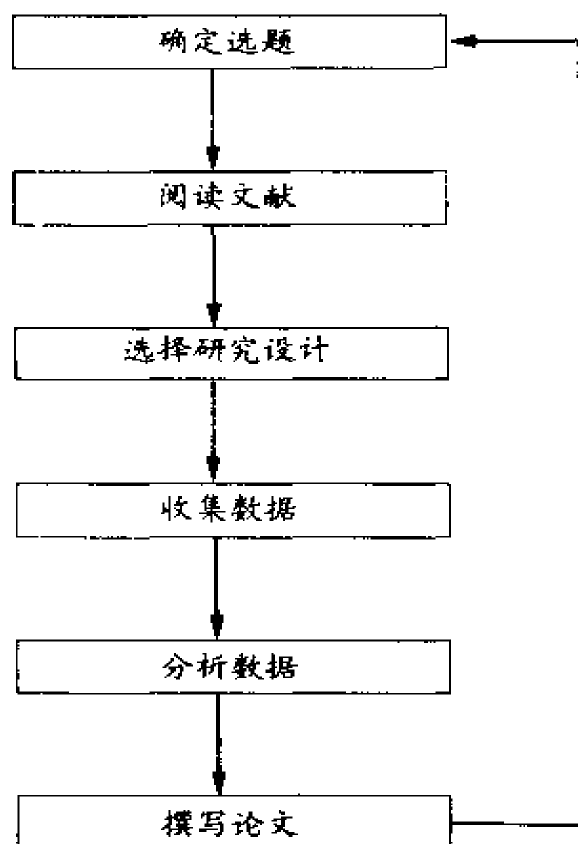


图 1.3：研究流程图

第三节 研究的分类

不同的研究可以根据研究目的和数据来源这两个参数进行分类（参见表 1.1）。依据不同的研究目的，研究可以分为基础理论研究和应用性研究；依据不同的数据来源，研究可分为文献研究和实证研究。

分类依据	研究类型
研究目的	基础理论研究和应用性研究
数据来源	文献研究和实证研究

表 1.1 研究的分类

1. 基础理论研究和应用性研究

基础理论研究的主要目的在于建构新的理论或检验已有的理论。应用性研究旨在解决教学中的实际问题。从事应用性研究的人往往与二语教学有着直接的联系，如：二语教师、二语教材编写者和二语测试设计者等。

基础理论研究

如前所述，基础理论研究的兴趣所在是理论问题，而不是生活中的实际问题。例如：Krashen 对“监控”理论的建构就可以看作理论研究。他的理论模式包括以下 5 个假说：

1）习得—学得假说；

2）自然顺序假说；

3）监控假说；

4）输入假说；

5）情感过滤假说。

上述理论最初并不是 Krashen 本人提出的。例如：输入假说最初是 MacNamara 在 1972 年提出来的（Krashen，1985），自然顺序假说最初是 Corder 在 1967 年提出来的。Krashen 通过研究、分析和综合相关文献，加上他本人的实证研究，建构了这一理论。

这类研究的例子在 *Language Learning*（《语言学习》）、*Applied Linguistics*（《应用语言学》）等国际学术刊物以及《现代外语》、《外语教学与研究》等国内学术刊物上俯拾皆是。我们来看几个例子：

例 1（Ye，1994）

本研究旨在测量文化背景知识对二语阅读理解的影响。研究者选择了两篇难度相当，有关不同文化的短文，一篇是关于中国的中秋节，另一篇是关于美国的感恩节。60 名非英语专业的学生参加了这项研究，他们在阅读结束后，接受了阅读理解测试。最后，研究者比较了两篇阅读理解的成绩。研究发现，学生对关于中秋节的短文理解得更好。

例 2（Wu，1998）

本研究旨在调查学生的语用知识与语言水平的相关性。90 名英语专业二年级的学生参加了语用能力测试。根据他们的英语总成绩，研究者把他们分为好、中、差 3 组。笔者运用统计方法，来比较这 3 组学生的语用能力。研究结果表明，学生的二语水平与他们的语用知识关系非常密切。

以上两个例子都想检验现行的理论假设。在例 1 中，Ye 以非英语专业的学生为研究对象，验证了“文化背景知识是影响阅读理解质量的一个重要因素”这一假说。在例 2 中，Wu 对语用知识与二语能力的密切关系进行了验证。尽管这两项研究的发现对二语教学均有价值，但是不能解决实际问题。

应用性研究

应用性研究旨在解决课堂教学或与教学相关的一些实际问题，这些研究的成果可以直接为同行试用和借鉴。请看下面两个例子：

例 1

本研究使用语言实验室进行口语教学，以探讨大班口语教学的新路子（Wen & Wu，1998）。参加这项研究的 29 名南京大学外国语学院英语专业二年级的学生来自同一个班级。这项研究为期 4 个月。学期结束时，要求学生用无记名方式填写“对运用语言实验室进行口语教学意见的问卷”，其内容涉及以下 4 个方面：1）对在语言实验室里上口语课的态度；2）比较在普通教室和在语言实验室里上口语课自己练习英语机会的多少；3）比较在普通教室和在语言实验室里说英语的紧张程度；4）比较在普通教室和在语言实验室里上口语课口语水平提高的速度。根据学生对问卷的回答，他们发现学生对在语言实验室里进行口语教学

所采取的态度是积极的。尽管在语言实验室里上口语课有一些不足之处，但利弊相比，还是利大于弊。

例 2

本研究旨在探讨在语言实验室里进行大规模口语测试的可行性。该项实验前后持续了 5 年，研究的内容涉及以下 3 个方面：1）测试的内容与难度；2）测试的实施；3）对磁带的评估。来自 60 所不同高校的 3,300名英语专业二年级的学生参加了这项实验。每次实验后，研究者都根据学生和教师对测试所反馈的意见进行修改。最终在 1995 年 5 月该项实验被确立为全国英语专业四级口语测试的形式（文秋芳，2000）。

以上两个例子都属于应用性研究，旨在解决实际问题。在例 1 中，Wen 和 Wu 研究如何更为有效地进行大班口语教学。在例 2 中，作者旨在设计能在全国各地进行的大规模口语测试。尽管这两项研究成果对实际教学、教学大纲设计以及测试有直接影响，但我们不能因此就说：从理论上讲，在语言实验室进行口语教学，或在语言实验室里测试口语比其他形式要更好。

研究目的：是彼此分离还是相互联系？

在上述的讨论中，Wen 根据研究目的列举了不同的例子说明基础理论研究与应用性研究的区别，这可能会给人一个印象：基础理论研究与应用性研究泾渭分明，互不关联。其实不然，基础理论研究与应用性研究的界限只是相对的。在实践中，研究往往带有双重或多重目的。基础理论研究具有实用价值，应用性研究也有理论价值。因此，基础理论研究与应用性研究只是各自目的的侧重点不同而已。

2. 文献研究和实证研究

根据资料来源，研究可分为文献研究和实证研究。文献研究的资料是文件、著作或杂志中的文章或者已经发表的、别人收集的数据，而实证研究的数据来自现实生活，是第一手信息。

文献研究

文献研究又称为图书馆内研究。例如教师要求硕士生完成一篇学期论文，题目为“选择两个你认为最重要的学习者因素来说明二语学习中的个人差异”。要想完成这篇学期论文，学生首先要到图书馆去查找相关资料阅读，在广泛阅读的基础上，确定两个学习者因素。设想某个学生选择了“动机”与“学习策略”这两个因素。他就需要查找并阅读含

有这两个变量的文章，其中有的是理论阐述，有的是别人就这两个变量所从事的实证研究。在阅读的基础上，他需要进行整理、归纳、综合，最后形成一个完整的、有条理的论文。这样的学期论文能够帮助读者系统了解现有文献对某个问题的理论纷争与实际的研究结果。好的学期论文还可以提出富有创造性的见解。

综述性文章是典型的文献研究。研究者回顾某个领域最近的研究，然后对已发表的成果进行总结、分析、评价或综合。虽然研究者回顾的材料并不是新的，但是一篇好的综述性文章同样具有真知灼见。如：Oxford 和 Crookall 在 1989 年发表的"语言学习策略研究：方法、发现及教学问题"就属于这一类文章。他们根据前人在这一领域使用的研究方法，对各种有关语言学习策略的实证研究进行描写和评价，最后，对以后这一问题的研究，提出了一些很有价值的建议。

实证研究

实证研究的所有资料和数据一般来说都是研究者亲自收集的。研究者既可能从二语学生那儿收集，也可能从二语教师那儿收集。实证研究可以是基础理论研究，也可以是应用性研究。请看下面几个例子：

例 1

为了调查善学者与不善学者在阅读策略上的差异，研究者对 60 名非英语专业的学生进行了问卷调查。这份问卷采用的是 5 分量表，即：1 = 这句话完全不适合我；2 = 这句话通常不适合我；3 = 这句话有时适合我；4 = 这句话通常适合我；5 = 这句话完全适合我。整个问卷由 25 个问题组成，选择其中一个数字填在每句话后的括号里。最后用统计分析方法对数据进行比较。

例 2

为了调查善学者与不善学者在阅读策略上的差异，研究者把 12 个学生平均分成两组，一组为善学者，另一组为不善学者。接着，研究者要求他们阅读一篇短文，阅读时用有声思维的方式把他们大脑里所想的报告出来。整个过程用录音机记录下来，然后再整理成文字材料。根据口头报告对阅读策略进行分类，最后再把这些类型加以比较。

例 3

为了研究英语专业二年级的学生在二语写作过程中使用母语的情况，研究者就在写作不同阶段是否使用母语这一问题对 50 名学生进行了问卷调查。问卷共 15 题，采用 3 分量表，即：从不—有时—通常。研

究者首先对问卷调查的结果进行统计分析，然后从50个学生中，分别选出8个使用母语最多的人和使用母语最少的人进行访谈。在访谈中，要求访谈对象详细描述二语的写作过程并解释在写作过程中怎样使用母语和为什么使用母语。最后，对访谈收集的材料进行分类，作为问卷调查的补充。

近年来，随着电脑技术的发展，国内外已建立了许多语料库。这些语料库为研究者提供了翔实的语料，其中既有书面语语料，又有口语语料。对于一个研究新手来说，使用现有语料库的数据进行研究非常方便，因为任何一个研究者在时间和资金有限的情况下，都很难大规模地收集数据。这样，就出现一个问题：使用语料库进行研究，究竟是实证研究还是文献研究呢？笔者认为，就语料库的数据而言，它们是别人收集的，然而，这些原始数据尚未有人如此分析过，因此，这类研究仍属实证研究。

通过上述讨论，笔者试图解释实证研究有别于文献研究。实际上，无论是实证研究还是文献研究都不能孤立地进行。文献研究依赖于实证研究，而实证研究又起源于文献研究。

3. 对研究生的要求

作为应用语言学研究生，有人可能要问："我们应该做哪一类的研究？"我们的回答是："4类研究都应该学会。就学位论文而言，应该是文献研究与实证研究的有机结合。就学期论文而言，一般是文献研究，但不排斥小型或微型的实证研究。值得一提的是，博士论文中的实证研究必须具有理论价值。

小结

"研究"可定义为：用系统方法探求问题的答案。任何一项好的研究都必须满足3个条件：1）问题要有价值、有新意、有可行性；2）研究步骤应是有理有据、清楚、明确，可以重复进行；3）答案具有效度。对研究过程的描述有"循环说"和"流程说"两种。"循环说"强调的是研究的动态方面，而"流程说"强调的是研究应该完成的6项任务。每个研究新手都应按部就班地完成任务，等到熟练后，可灵活调整顺序。研究就其目的而言，可分为基础理论研究和应用性研究；就其数据

来源而言，可分为文献研究和实证研究。这 4 种研究，研究生都要掌握。而在应用语言学方向的硕士和博士论文中，文献研究与实证研究是论文的主体。

讨论题

1. 关于"研究"，这本书中的定义与你的定义以及你在其他书上看到的定义有什么不同？
2. 在近期出版的国际杂志上找一篇研究论文，和同学讨论下列几个问题：
   1）这篇论文研究的问题是什么？
   2）这篇论文采用了什么研究方法？
   3）研究问题的答案是什么？
3. 问题、系统方法和答案这 3 个研究要素之间的关系是什么？
4. 比较"循环说"与"流程说"，找出它们的共同点与不同点。
5. 研究者怎样才能进行有效的研究？
6. 如何给不同类型的研究分类？
7. 举例说明什么是基础理论研究，什么是应用性研究。
8. 实证研究在哪些方面需要文献研究的帮助？
9. 如何进行文献研究？

第二章　基本概念

本章介绍这本书中反复出现的一些基本概念。理解这些基本概念是理解其他章节的基础。这一章要讨论的概念包括：假设、理论、理论模式、总体、样本、变量、测量方法、理论定义和操作定义。

第一节　假设、理论及理论模式

不管从事哪类研究，都不可避免地要使用假设、理论及理论模式这些概念。若没有这些概念，人们会对研究的科学性表示怀疑。当然，仅仅靠使用这些概念并不能保证研究的科学性。

1. 假设（hypothesis）

如前所述，所有的研究都旨在探求问题的答案。有时，研究者在研究之前，就对问题的答案有了某种设想，研究的目的就是用证据来证实或推翻自己的设想。这种对答案的设想称作为“假设”。准确地说，假设应定义为对两个或两个以上变量之间关系的预测。这样的假设是建立在前人研究成果的基础上或是建立在已确立的理论基础上，而不是凭空想像的主观臆断。

比如说，我们的研究问题是：“在英语学习过程中，使用母语的程度与学习成绩有什么关系?”从逻辑的角度讲，它们之间的关系可能有4种假设：

1）零假设亦称为虚无假设或原假设，用 H_0 来表示，指使用母语程度与英语学习成绩之间没有关系的假设；

2）正向假设，用 H_1 来表示，指使用母语程度与英语学习成绩之间有正向关系的假设；

3）负向假设，用 H_2 来表示，指使用母语程度与英语学习成绩之间

有负向关系的假设；

4）无向假设，用 H_3 来表示，指使用母语程度与英语学习成绩之间尚未确定是正向还是负向关系的假设。

“假设”可分为零假设和备择假设，备择假设又可分为正向假设、负向假设和无向假设（参见图 2.1)。

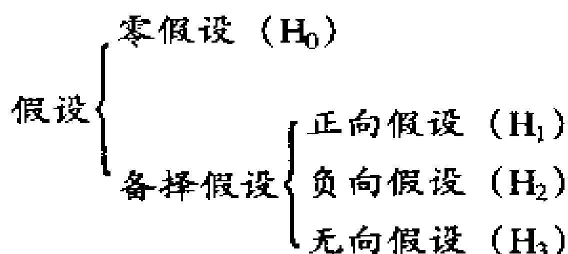


图 2.1：假设类型

根据前人研究的大量证据，预测两个变量之间的关系是正向的还是负向的，这样的假设叫作有向假设。有时，前人研究的结果不尽一致，有些甚至彼此矛盾，这时无法确定其方向，这时形成的假设叫无向假设。就上述研究问题而言，如果我们认为使用母语程度与英语学习成绩有关系，但不知道是什么样的关系，这就是无向假设。如果我们根本不知道它们之间是否有关系，这时就提出零假设，即两个变量之间没有关系。

零假设在研究中有着特殊的意义。严格说来，一切检验假设的研究都开始于零假设。其原因很简单。在实际生活中，没有任何统计工具能够用数据直接证明有向假设和无向假设，统计工具所能做的就是接受或者推翻零假设。尽管有的研究表面上看起来不是从零假设开始，但研究背后的逻辑推理仍旧与零假设有关。这就好比我们在法庭上难以用证据直接证明某人清白一样。法官必须从零假设开始，即某人不是罪犯。只要有人在法庭上拿出证据足以证明某人犯了法，法官就会推翻零假设。如果证据不充分，法官就只能接受零假设（参见第八章)。

有些研究事先形成了假设，然后再进行验证；有些研究则事先没有假设，问题的答案在研究过程中才渐渐形成，其答案可以成为以后研究有待检验的假设。例如：文秋芳、郭纯洁（1998）研究了学生在英语看图作文中使用母语的情况。他们在开始研究时，并不知道学生在英语看图作文中如何使用母语，因此不可能形成假设。他们要求学生一边作

文，一边说出内心的思维活动。研究者在分析学生复杂的内心活动过程中，才逐渐了解了母语在英语看图作文中的不同作用。此项研究的结果就是假设，这些假设需要在其他研究中进一步验证。

总之，假设可能在研究的不同阶段形成，然而，它的形成对于任何研究来说都是至关重要的。研究可以从假设出发，研究过程就是一个验证假设的过程；研究也可以以假设结束，研究过程就是产生新假设的过程。

2. 理论（theory）

理论由一组互相关联的假设组成*。这一组假设对所研究的对象形成系统、完整的阐述。例如，Krashen（1985）提出的二语习得理论包括5个假设（PP.1－3）：

1）习得—学得假说

习得与学得是两个互不关联的培养外语能力的途径。习得类似儿童学习母语，在潜意识中进行，能够获得自动化的技能；而学得是有意识的行为，只有助于语言知识的掌握。

2）自然顺序假说

学习者习得语言规则的顺序是可以预测的，但它既不依赖于感知形式的复杂性，也不依赖于教学的顺序。

3）监控假说

通过有意识学得的知识，可以用于监控语言行为，但这里有两个前提：①学习者有意关注准确性；②学习者必须了解相关规则。

4）输入假说

如果学习者大量接触稍微超出其现有语言水平的可理解性的输入，他们便可以习得二语。

5）情感过滤假说

情感过滤器好比一个可以上升或下降的心理障碍物。当它上升时，外部输入就不能进入到长期记忆中去；当它下降时，外部输入进入长期记忆就没有障碍。

以上5个假说紧密联系，形成Krashen的二语习得理论：输入假设

* 需要指出的是，有的研究者对“假设”与“理论”这两个术语不做进一步区分，两者经常混用，但我们认为还是区分的好。

说是习得的必要条件；学得的知识只能用于监控语言行为；习得语言的顺序不受外界因素影响；习得语言时情感过滤器不起阻碍作用。归根结底，Krashen 的理论强调潜意识的自然习得，忽视有意识的刻意学习。自 Krashen 的理论提出后，许多追随者试图通过研究找到实证依据，但在所得到的研究结果中，有的支持 Krashen 的理论，有的却与之大相径庭；在验证 Krashen 理论的过程中又涌现出新理论，例如 Long（1980）的互动交际理论和 Swain（1985）的输出理论。这就是说，一种新理论的提出能够引发大量验证性的实证研究，在验证性的研究中，又能产生更新的理论，如此循环，推陈出新，科学就能不断前进。

再如：文秋芳、郭纯洁（1998）在研究二语学习者看图作文中使用母语的情况时，根据学生的口头报告，提出一个理论。这一理论包括以下一些假说：

1）学生接受了图画刺激后，可以有 3 种选择开始英语写作：①图画⇒外语；②图画⇒转换中介；③图画⇒内容生成中介。学生无论选择哪一种途径，几乎都避免不了使用母语。

2）母语在外语看图作文过程中一般作为转换中介、内容生成中介、形式检索中介、内容验证中介和程序管理中介，它们在外语句子（从句）形成的过程中交替出现。

3）从总体上看，外语写作能力高的学生对母语依赖性小。其中高能力组较多使用中文作为转换中介和内容验证中介；而低能力组较多用中文作为内容生成中介、形式检索中介和程序管理中介。就使用中文的平均长度而言，高能力组用于转换中介和内容验证中介的中文平均长度偏长，而低能力组用于内容生成中介和形式检索中介的中文平均长度偏长。

Wang（2000）最近对上述理论进行了验证性研究。她发现母语功能的种类远远超过上述五类。

3. 理论模式（model）

尽管在文献中，有人将“理论模式”和“理论”这两个术语混用，但也有许多研究者赞成将这两个术语区别开来。“理论模式”指的是着重用图形来描述的某个有系统的结论，而“理论”指的是着重用语言来描述的某个有系统的结论。例如：Krashen 的理论也可用下列图形来表述，这时就称为理论模式（见图 2.2）。

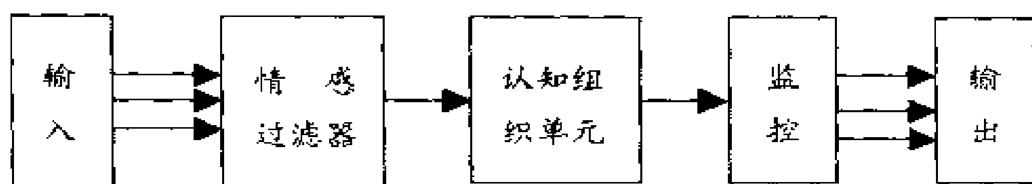


图 2.2：二语习得理论模式

显然，图形表述比语言描述更形象、更生动。但仅有图形还不够，在描写变量之间的相互关系时，语言比图形更加清晰、更加明确。因此，笔者建议无论在写硕士论文还是写博士论文时，最好是既有图形表述，又有语言描述，这样会便于读者理解。

第二节 研究对象总体与样本

1. 研究对象总体界定

研究对象总体是指所要研究的全部人或事物的总和。它可以是人、课本、试卷等。在大部分情况下总体用于指人。例如：如果研究者想测量南京中学生的英语水平，那么研究对象总体就是南京所有的中学生；如果研究者对过去 10 年高考英语试题的难度进行评估，那么研究对象总体就是 10 年的高考英语试题。

样本是研究对象总体中具有代表性的一部分。样本的抽取通常采用随机程序，即在抽样过程中，总体中的每个个体被抽中的机会完全相同。经过随机程序抽取的样本称为随机样本。第六章将详细介绍随机抽样技术。

2. 研究对象总体与样本的关系

如果样本是随机抽取的，从统计的角度，我们就可以认定样本是研究对象总体的一个缩影。通过研究样本，研究者最终能推测出整个研究对象总体的情况。

如图 2.3 所示，首先从研究对象总体中抽出样本，然后对样本进行研究。一旦有了样本研究的结果，研究者就可以从样本结果中推测出研究对象总体的情况。这样推测的准确性有多高，取决于样本在多大程度

上能够代表研究对象的总体。

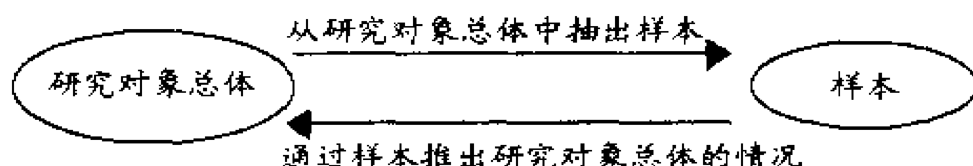


图 2.3：研究对象总体与样本（Punch，1998:106）

第三节 变 量

变量是指数值可以变化的量，例如学生的成绩。为了研究问题的方便，我们常常把所要研究的要素，人为转化成一个数值，比如：性别、婚姻状况、受教育程度等。通常，人们将男性设定为 1，将女性设定为 0 或 2。有些变量的变化极其有限。例如：性别的变化只有 2 种：男性或女性。婚姻状态的变化有 4 种：未婚、已婚、离婚和丧偶。教育程度的变化可能有 4 种：初等教育、中等教育、高等教育和研究生教育。然而，有些变量的变化范围比较大。例如：学生成绩的变化可以从 0 到 100。身高和体重的变化比考试成绩的变化还要多。根据变量的作用，我们把变量分为自变量（independent variable）、因变量（dependent variable）、调节变量（moderator variable）、控制变量（control variable）和干扰变量（intervening variable）。下面我们逐一讨论。

1. 自变量与因变量

自变量与因变量关系非常密切，互相依存。没有自变量就没有因变量。同样，没有因变量也就没有自变量。

自变量与因变量可在 2 种情况下使用，第一种情况，自变量是一个刺激变量（stimulus variable）或输入变量（input）。因变量是一个反应变量（response variable）或输出变量（output）。例如：研究者研究 X 和 Y 两个变量之间的因果关系。如果它们之间的关系可解释为："如果 X 变大或变小，Y 会怎么样？"那么，X 就是一个自变量，Y 就是一个因变量。

第二种情况，自变量是用来预测另一个变量，而因变量是个可以被

自变量所预测的变量。它们之间的关系不是因果关系，而是预测与被预测的关系。例如，某研究者进行一项相关研究，旨在弄清性别在多大程度上可以预测学习者的二语成绩。在这项研究中，性别是一个自变量，二语成绩是一个因变量。更为确切地说，性别是一个预测变量，二语成绩是一个被预测变量。

一言以蔽之，在第一种情况中，自变量是个动因，引起因变量的变化；在第二种情况中，自变量是个用来预测另外一个变量的变量。有的研究只有一个自变量和一个因变量；有的研究涉及多个自变量和一个因变量；也有的研究涉及一个自变量和多个因变量。下面我们举些例子，来看看哪些是自变量，哪些是因变量。

例 1

假设：有两种二语写作教学法，一种要求学生把同一篇作文修改多次，简称为“反复修改法”。另一种要求学生多写一些不同的文章，简称为“多篇作文练习法”。该项研究认为反复修改法比多篇作文练习法对提高学生的写作能力更有效。

自变量：写作教学方法

方法 1：反复修改法

方法 2：多篇作文练习法

因变量：二语写作能力的提高

在例 1 中，由于不同的写作教学方法能够导致写作能力提高的程度有所不同，因此写作教学方法是一个自变量，而二语写作能力的提高是一个因变量。

例 2

假设：频繁使用多种学习策略的人比使用学习策略种类少且频率低的人英语学习成绩要好。

自变量：使用学习策略的种类与频率

因变量：英语学习成绩

在例 2 中，研究者假设使用学习策略的种类与频率能够影响二语学习成绩。使用学习策略的种类越多，且频率越高，他们取得的学习成绩就越好。也就是说，学习策略的使用与英语学习结果之间存在因果关系。因此使用学习策略的种类与频率是自变量，英语学习成绩是因变量。

例 3

假设：学生对教育的态度可部分预测他们对“好”老师的看法。

自变量：学生对教育的态度

因变量：学生对“好”老师的看法

在例3中，研究者假设了解学生对教育的态度可以预测学生对“好”老师的看法。因此，前者是一个自变量，后者是一个因变量。

例4

假设：父母的教育程度能够预测学生学习成绩的好坏。

自变量：父母受教育程度

因变量：学生的学习成绩

在例4中，研究者假设父母的教育程度与子女的学习成绩存在着密切的关系。这就是说，父母的教育程度越高，子女的学习成绩可能越好。很显然，父母受教育程度与学生的学习成绩之间关系不是因果关系。因此，父母受教育程度是预测变量，学生的学习成绩是被预测变量。

2. 调节变量

调节变量是一个可以调节自变量与因变量之间关系的变量。它有时又被称作为第二自变量。“调节”这个词表明了为什么要把第二自变量专门挑出来进行研究。如果研究者研究自变量X对因变量Y的影响，但怀疑X和Y之间的关系可能由于因素Z的影响而发生改变，因此Z就作为调节变量来研究。

我们来看两个例子。第一个例子：研究者在教英语时想看视觉刺激（使用图画）和听觉刺激（使用磁带）哪种方法更为有效。首先，她随便把学生分成两个班，一个班用视觉刺激，另一个班用听觉刺激。经过3个月的实验，所有的学生参加同一测验，结果，两个班的成绩差不多。这时，研究者想到有些学生喜欢用眼睛接受信息，有些学生喜欢用耳朵接受信息，很可能学生所偏向的学习风格会影响到学习成绩。因此，她设计了问卷来测量学生所偏爱的学习风格。然后，她把偏爱视觉刺激的学生和偏爱听觉刺激的学生分开，原来模糊不清的图画，一下子豁然开朗。如果把这两种方法与偏爱这些学习风格的人搭配起来，那么这两种方法都非常有效。在这种情况下，学生所偏爱的学习风格可以被看作调节教学方法（自变量）与方法的有效性（因变量）的调节变量。

第二个例子是研究考试环境（自变量）与考试成绩（因变量）之间的关系。假定研究者提供两种考试环境。一是以考生为目标（把姓名写在试卷上，我们测量的就是你）；另一是以试题为目标（不把姓名写在

试卷上，我们测量的是试题的难度）。考生的焦虑程度可看作成调节变量。结果发现焦虑程度高的人在以试题为目标的考试中考得比较好；焦虑程度低的人在以考生为目标的考试中考得比较好。这就是说，自变量、调节变量和因变量之间有相互作用。二语教学的情况非常复杂，在研究中最好至少要包含一个调节变量。常常由于研究者没有把一些至关重要的调节变量挑选和测量出来，使得自变量与因变量之间的关系显露得不清楚。我们再来看一些例子。

例 1

假设：男性研究者比女性研究者更为有效地获取所有研究对象的信息，而女性研究者比男性研究者更为有效地获取男性研究对象的信息。

自变量：研究者的性别

调节变量：研究对象的性别

因变量：获取研究对象信息的多少

例 2

假设：在学习英语的过程中，使用母语对学生的学习成绩有消极影响。该消极影响对成人比对儿童更为明显。

自变量：使用母语

调节变量：年龄

因变量：学习成绩

例 3

假设：管理策略对英语学习成绩有影响，且对低水平学习者的影响更大。

自变量：管理策略

调节变量：英语水平

因变量：学习成绩

3. 控制变量

实际工作中，我们往往不可能同时对所有变量进行研究。当我们使其中某些变量保持中立，确保这些中立的变量不影响自变量与因变量的关系。这些保持中立的变量就称为“控制变量”。确定控制变量的目的就是控制这些变量可能会对研究结果有影响。例如：在一项关于课外阅读对培养阅读理解能力的影响的研究中，研究者有意地选择了四级英语80分以上的学生作为研究对象。这样，研究对象的总体水平就是控制变

量。在另一项研究中，研究者想弄清测试策略的训练在多大程度上可以影响测试成绩。所有的研究对象都是女生，这样，性别就是控制变量。

尽管某些变量偶尔也充当调节变量，但它们在大部分情况下都是控制变量。例如：性别、智力和学习者从前的学习情况。研究者在设计研究时，总需要决定哪些变量需要研究，哪些变量需要控制。

4. 干扰变量

干扰变量是指在研究中任何一个没有测量出来，但和自变量与调节变量一起对因变量产生影响的变量。研究中有两类干扰变量。第一类是指研究者想测量，但很难测量的变量。第二类是指可以测量，但研究者因时间、资源有限或研究经验不足而忽略的变量。一般说来，这两类干扰变量常常同时出现。例如：你想进行一项研究旨在搞清楚阅读量（自变量）与词汇量（因变量）的关系。在这两种变量之间，还有另外一些变量在起作用，如：阅读动机、已掌握的词汇量、记忆单词的策略、个体记忆能力。

再如：你想比较是利用闭路电视上英语课的效果好，还是现场听老师上课效果好。这项研究中的自变量是上课方式，因变量是学习效果。经过 3 个月的实验，结果发现在现场听老师上课比通过闭路电视上课效果要好。很显然，教学方式对学习结果有影响。如果你问为什么一种教学方式比另一种教学方式好，回答可能是现场听老师上课可以给学生提供参与课堂讨论的机会，而闭路电视教学只能单向交流。因此，学生的参与量就是一个干扰变量，它可能会对学生的学习结果产生影响。实际上，你可以列举出一连串的干扰变量，如：重新播放电视的机会、学生对使用闭路电视不熟悉，等等。

一个好的研究者必须能够在研究中识别干扰变量。知道什么是干扰变量会有助于解释自变量引起因变量变化的原因。不仅如此，弄清干扰变量的影响会使研究者在解释研究结果时谨慎从事，而不至于把话说得太满，甚至解释错了。

5. 变量之间的关系

上述 5 种变量之间的关系怎么样？下面根据 Brown（1988：13）的图（参见图 2.4）说明了它们之间的关系。自变量与因变量之间的关系要么是因果关系，要么是相关关系。调节变量又称为第二自变量，可改

变自变量与因变量之间的关系。控制变量在研究中保持恒定，即没有变化，目的是更好地研究自变量对因变量的影响。干扰变量不是直接可以观察到的，因此，放在虚线箭头里。干扰变量越多，就越难说清楚两个变量之间的因果关系。

一个变量的功能并不是这个变量本身所固有的。因此，变量在图中的位置不是一成不变的，它随着研究的变化而变化。有时在同一研究内，在不同的研究问题内变量的功能也可能不同。例如：动机这一变量在这个研究里是自变量，到了另一个研究里就变成因变量、调节变量、控制变量或干扰变量。研究者必须要确定某个变量在不同研究问题中的不同功能。

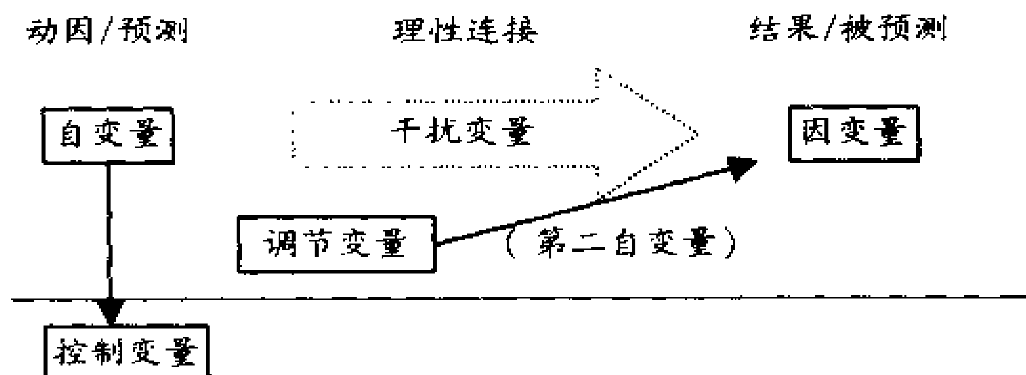


图 2.4：不同变量之间的关系

第四节 测量方式

变量除了功能上的差异以外，还有测量方式上的差异。一般说来，应用语言学有 3 种不同的测量方式：称名测量、等级测量和等距测量。与之相对应的有 3 种变量：称名变量、等级变量和等距变量。下面我们来逐一讨论这 3 种测量方式。

1. 称名测量

顾名思义，称名测量是用来给事物命名或给事物分类。一个称名变量包含几个类别，这些类别彼此相互排斥。例如性别这个称名变量包含两个类别：男性和女性。一个人就其性别而言，要么是男，要么是女，不可能既是男，又是女。换句话说，每个类别都是独一无二的。我们可

以根据母语、性别或国籍把研究对象进行分类。尽管称名测量非常简单，但许多极为有趣的变量都属于这一类，常见的称名变量有性别、职业和母语。

就其实质而言，称名测量不是一种真正意义上的测量，因为它不涉及数学上的数量概念。为了统计的方便，我们常常把这些不同的类别转换为数字，例如：把男性标为 1，女性标为 2；也可以把女性标为 1，男性标为 2。通过统计，我们很快就能得到样本中的男女百分比。但这里的 2 没有一般数字上的意义，它不是 1 的 2 倍，也不能进行加减乘除。哪个数字代表哪一类型是随意的，只要每一类有个数字就行。从这个意义上说，称名测量就是给某个变量内部的不同种类冠以一个名称，以便区分。下面我们来看几个例子。

例 1：性别

1 = 男性

2 = 女性

例 2：婚姻状况

1 = 未婚

2 = 已婚

3 = 离婚

4 = 丧偶

例 3：母语

1 = 汉语

2 = 俄语

3 = 英语

4 = 日语

5 = 朝鲜语

6 = 瑞典语

例 4：国籍

1 = 美国

2 = 法国

3 = 德国

4 = 英国

5 = 澳大利亚

6 = 加拿大

2. 等级测量

等级测量除了给人或事物命名以外，还提供他们拥有特征的相对量级。例如：在英语演讲比赛中，最终选出6名最好的选手，分别为第一名、第二名、第三名、第四名、第五名和第六名。这样一个排列顺序就是一个等级测量，表明谁比谁好。然而，第一名和第二名之间的差距，以及第二名与第三名之间的差距，我们不能准确地测量出来。再如：老师在评价学生作文时，没有给出精确的分数，而是给它们5个等级，即A、B、C、D和E。很明显，得A的学生比得B的学生作文写得好。可A与B或B与C之间到底有多少差异呢？这正是等级测量所缺少的。

在数据编码过程中，人们常用阿拉伯数字代表等级。例如：用5代表A，用4代表B。用数字表示等级并不能把两个等级之间的差距准确地测量出来。如果想比较两个等级变量，我们只能说$1<2<3$或$3>2>1$。在英语口语考试中，我们可以说张比李考得好，但没有朱考得好。

3. 等距测量

等距测量除了能完成称名测量和等级测量所作的一切以外，还能保证每两个连续点之间的距离是相等的。例如：1分钟到2分钟与4分钟到5分钟的时间是相同的。同样，2米到3米与4米到5米的距离是相同的。因此，在测量时间、长度和距离等时，经常用到等距测量。

在学校里，考试使用的百分制属于等距测量，每两点之间的距离是等值的，都是1分。如果有两个学生，一个考了100分，另一个考了95分，我们可以说一个学生比另一个学生考得好，因为他多考了5分。这5分当中，每两个分值之间的差距完全相等。百分制就是建立在这样一个假设的基础上，即1分到2分与5分到6分，或99分到100分的距离相同。

然而，上述观点似乎过于简单化。考试项目通常难度不一，因此，不同项目所得的分数不一定等值。但在考试中，难度差异一般被忽略不计，我们把这样的测量方式仍旧看作是等距测量。

等级测量与等距测量的差别可在图2.5上体现出来，这个图是根据Bachman（1990：28）的图修改而成。例如：考第一名的学生得了95分，考第二名的学生得了93.5分，考第三名的学生得了90分。第一名与第二名之间的差距是1.5分，而第二名与第三名的差距是3.5分。由此观之，

第一名和第二名之间的差距，与第二名和第三名之间的差距是不等的。与之相对应的等距测量，每两个分值之间的差距是相等的，都是 1 分。

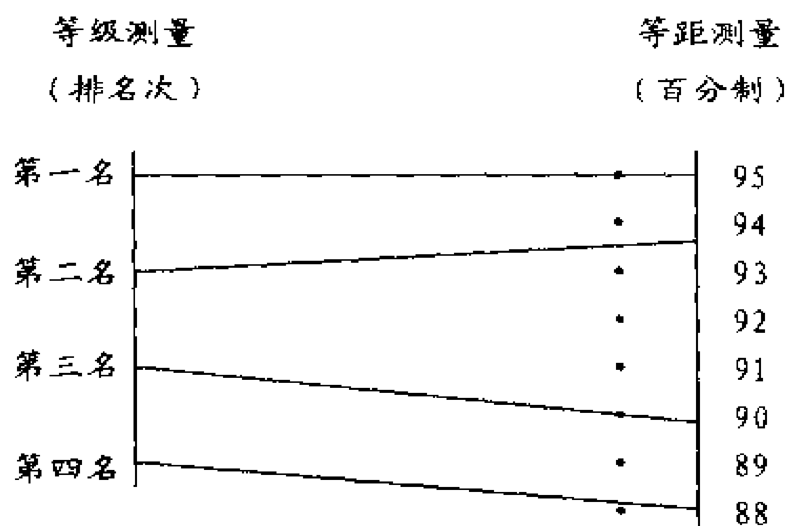


图 2.5：等级测量与等距测量的比较

比较理想的是，在用等距测量测量一个变量时，变量之间的差距是相等的。一般情况下，这一点比较容易做到，如时间和重量。但在另一些情况下，就不能保证。例如：年龄是研究的一个变量，你可能认为一年就是一年。然而，对于二语学习者来说，每两年之间的差异并不相等。很明显，在 10 岁至 20 岁之间，每两年的差异或许要比在 60 岁至 70 岁之间每两年的差异要小得多。

研究中，我们经常采用 3 分量表或者 5 分量表，通过问卷调查学生对某个问题的看法或做法。严格说来，这样的测量只能是等级测量。但这些研究往往把这样的等级测量当作等距测量来对待（Boyle，1970；Labovitz，1971）。例如：目前我们还没有任何工具可以准确地测量出人们对学习二语的兴趣。问卷中采用不感兴趣、感兴趣和非常感兴趣三个不同类别让被调查者选择。接着给每一类一个数字，如：不感兴趣 = 1，感兴趣 = 2，非常感兴趣 = 3。最后把 1 – 2 – 3 作为等距变量处理，计算出平均数。从理论上讲，这种测量问题颇多，首先，我们无法得知不感兴趣、感兴趣与非常感兴趣之间的距离，其次每两个连续点之间的距离难以相等。然而，社会科学中的许多研究都把这样的问卷调查看作等距测量，为的是便于统计。目前这一做法已广为人们接受（Rudestam & Newton，1992）。

4. 三种测量之间的关系

3种测量中，就准确性而言，称名测量处在最低级，最不准确；等距测量处在最高级，最准确；等级测量处在两者之间，比称名测量准确，但不及等距测量。称名测量只表示类别；等级测量不仅表示类别，而且表示类别的顺序；等距测量除了表示类别和类别的顺序还表示顺序排列之间的距离。根据 Brown（1988：23）的测量方式，我们把3种测量总结如下（参见表2.1）。

测量方式			类别	顺序	距离
最高级	↑	等距测量	+	+	+
		等级测量	+	+	−
最低级		称名测量	+	−	−

表2.1　测量方式及其特点

较高级的测量可转变为较低级的测量，但较低级的测量不能转变为较高级的测量。以学生的水平测试为例，我们可以把百分制变成5个等级，但不可能把5个等级转变为百分制。

虽说用什么方式测量变量取决于研究者，但研究的一般规则是“尽可能用最高级的测量方式来测量。这就是说，如果能用等距测量，决不要用等级测量”（Bernard，1994：35）。

第五节　如何将理论定义变为操作定义？

在本节中，我们先讨论理论定义和操作定义的差别，然后讨论如何将理论定义变为操作定义。

1. 理论定义和操作定义

一个变量需要从理论和操作两个方面进行界定。理论定义提供人们理解某个变量的概念特征。操作定义提供人们测量某个变量的具体方法。下面的例子可说明理论定义与操作定义的差别。

我们先来看日常生活中的一个例子。你可能听别人说“老张的房子

很宽敞。”可这里的“宽敞”是什么意思？《朗文英语和文化词典》把“宽敞”界定为“有许多空间”（1993：1271）。根据这一定义，人们很难确定某人的房子是否宽敞。在研究住房问题的时候，研究者把“宽敞”定义为人均面积在 15 平方米以上。根据研究者这一定义，我们就可以确定某人房子是否宽敞。由此观之，词典上的定义是理论定义，研究者的定义是操作定义。

我们再看一个第二语言研究方面的例子。Gardner（1985）把动机定义为“努力 + 实现学习目标的愿望 + 对学习语言的积极态度”（p.10）。也就是说，动机包括 3 个部分：态度、愿望和努力。这是动机的理论定义。Gardner 难以用这个定义去测量每个学习者的动机。于是他根据动机的 3 个部分分别设计了 3 个问卷让学习者去回答。每个学习者的动机就是对这 3 个问卷的反应。这里 3 份问卷就成了动机的操作定义。

从上面例子可以看出，虽然理论定义有助于我们理解变量的概念特征，但用途极为有限，我们不能借助理论定义进行测量。没有测量，我们就不能直接进行比较。有了操作定义，研究者不仅可以测量某个变量，而且可以讨论、评价其合理性。

在应用语言学和其他社会科学领域里，理论定义常常出现在文献回顾中，操作定义常常出现在研究方法的描述中。在给变量下理论定义时，必须注意到由于研究者的观点不同、又由于理论的发展，在不同文献中，同一术语的理论定义不一定相同。碰到这样的情况，研究者必须选择适合自己研究的理论定义，或者根据自己的研究，对现有的理论定义稍加修改，并说明其理由。以二语水平为例，二语水平曾一度被定义为语法、词汇和语音 3 方面知识的总和；20 世纪五六十年代时被定义为听、说、读、写四种技能的总和（Lado，1961）；70 年代时被定义为交际能力（Hymes，1972）；八九十年代时该定义进一步扩大，又增加了策略能力（Bachman，1990）。如果某人想研究学生的二语能力，就会面临数种选择。需要指出的是，选择的理论定义会直接影响其操作定义。比如说，如果把二语水平看成为 4 种技能，那么二语水平的操作定义必须相应地与 4 种技能有关。换句话说，在一项研究中，同一术语的理论定义与操作定义必须一致且逻辑上相关。同样，如果认为二语水平是由语言能力、语用能力和策略能力组成，而所选用的测试是托福，那理论定义和操作定义就不一致，因为托福不能测量语用能力和策略能力。

然而，理论定义与操作定义的一致并不意味着操作定义必须测量理

论定义中所描述的所有内容。人们常常因为测量工具的缺陷、时间的限制以及人力、物力和财力的不足，实际所测量的通常少于理论定义所界定的内容。换句话说，操作定义肯定不能百分之百地体现理论定义。例如：没有哪一个英语考试可以体现二语水平理论定义的各个方面。即使理论定义相同，不同的研究中使用的操作定义也可能不一样。例如：50年代时，有些研究者的语言能力的操作定义是 Carroll 和 Sapon 设计的“现代语言能力测试”（MLAT）的成绩。60年代时，另一些研究者的语言能力的操作定义是 Pimsleur 设计的“语言能力测试”（LAB）的成绩。尽管 MLAT 和 LAB 这两种测试都声称能测量语言能力，但准确地说，它们实际所测量的不是同一方面。Skehan（1989）指出“与 MLAT 相比，LAB 强调语言学习的推理能力和听的能力。……与 MLAT 相比，LAB 所缺少的是语法敏感性的测试以及有效记忆范围的测试”（p.29）。

从逻辑关系上说，我们应该首先确定变量的理论定义，然后再是变量的操作定义。在实践中，这一过程往往不是线性的。例如：也许你最初按照目前最全面的观点：二语能力包括语言能力、语用能力和策略能力。后来，你可能发现没有同时测量3种能力的权威测试。如果你坚持这一理论定义，你就无法使之操作化。因此，你必须修改理论定义。

2. 操作定义

变量的操作定义要明确说明如何测量理论定义所界定的内容。这一定义的特点是：清楚、明确。下面是 Wen（1993）在研究学习者可控因素与学习结果关系时，所给出的一些操作定义：

1）二语水平

英语专业四级考试的分数。

2）努力程度

研究对象报告的一周之内课后学习英语的时间。

3）学习目的

对问卷调查项目的回答，诸如：你学习英语的原因是对其语言和文化感兴趣还是渴望未来找到一份好工作，等等。

4）管理策略

对问卷调查项目的回答，诸如：你在计划、确定目标和评估学习过程中使用的策略以及学习者的个人特点。

如果我们能按照统一的操作定义测量某些变量，那么我们就可以比

较来自不同大学的学生在这些变量上的差异。例如：我们可以通过比较，看看甲大学的学生是否比乙大学的学生在英语专业四级考试中考得好；我们还可以通过比较，看看甲大学的学生是否比乙大学的学生每周课外花在学英语上的时间要多。

如何给一个变量下操作定义？首先要研究相应的理论定义，然后弄清实证研究中要测量理论定义中的哪几个方面。例如：在 Wen（1993）的研究中，“努力程度”这个变量的理论定义是课内与课外花在二语学习上的时间数量与质量。换句话说，这一理论定义包括两个部分：时间的数量和质量。很明显，时间的数量和质量对英语学习结果都非常重要。然而，时间的质量很难测量，再说，参加这项研究的研究对象都是英语专业大二的学生，他们每周的英语课时数差不多。因此，文秋芳给“努力”下的操作定义是“研究对象自我报告课外花在英语学习上的小时数”。很显然，这里的操作定义没有测量时间的质量，也没有测量英语教学的课时数。

操作定义的质量对研究非常重要。如果操作定义下得不好，所有用操作定义测量而得出的结论都不可靠。例如：如果把二语水平定义为学生参加某次课堂考试的成绩，这里的效度和信度都无法检验，因此此类的操作定义就不能用。当然，这样的问题往往在研究初期就会被导师或其他研究者发现。

小结

本章讨论的基本概念包括假设、理论、理论模式、研究对象总体、样本、变量、测量方式、理论定义和操作定义。假设是对研究问题答案设想的陈述。有的假设产生于研究之前，有的假设产生于研究之后。假设的真伪可以在研究中进行检验。零假设是指研究中的两个或两个以上变量之间没有关系的假设。备择假设包括有向假设和无向假设。理论由一组互相关联的假设组成，理论可以用文字来表述，也可以用图形来描绘。为了区别这两种表达方式，用语言表达的叫理论，用图形来描绘的叫做理论模式。总体是研究者所要研究对象的全体，样本是研究对象总体中具有代表性的一部分。从样本中获得的结果常常用来推断总体的情况。变量是指其数值因人、因物、因时而改变的量。变量就其作用而言，可分为自变量、因变量、调节变量、控制变量和干扰变量；变量按

测量方式划分，可分为称名变量、等级变量和等距变量。研究中所有的变量都要进行理论和操作界定。理论定义帮助人们弄清变量的概念特征，而操作定义帮助人们具体测量某个变量。

讨论题

1. 假设有哪些类别？从国际学术刊物上发表的论文中找出各种假设。
2. 理论与理论模式有何区别？请举例说明。
3. 什么是总体？什么是样本？总体与样本有什么关系？请举例说明。
4. 什么是自变量、因变量、控制变量、调节变量和干扰变量？在国际学术刊物上找一篇实证性研究论文，并指出文中不同类别的变量。
5. 理论定义与操作定义有什么区别？如何给一个变量下操作定义？
6. 从发表的论文中找出5个变量的理论定义和操作定义，然后与他人讨论。
7. 如何根据不同的测量方式来测量一个变量？从国际学术刊物上选一篇研究论文来说明不同的测量方式。
8. 从国际学术刊物上选择一篇适当的研究论文，从头到尾仔细阅读，然后列出这篇论文的基本概念，并向他人解释这些概念。

第二部分　研究方法

第二部分共分 10 章，分别叙述与讨论研究过程中各阶段应承担的任务与常见问题及解决问题的方法。通过这一部分的学习，应能达到下列目标：

- 了解研究的基本程序；
- 承担各研究阶段的具体任务；
- 掌握常见问题或困难的解决办法。

第三章　提出研究问题

要找到一个好的研究问题通常耗时费力。研究者不仅需要阅读大量的文献，而且需要与其他研究者交流研讨。在这个过程中，研究者必须反复考虑选题是否具有理论意义和实践意义，是否具有新意，是否切实可行。本章将描述提出研究问题的一般程序，常见困难，以及解决困难的建议。

第一节　提出问题的步骤

研究问题的提出一般需要经过以下几个步骤，如图 3.1 所示：

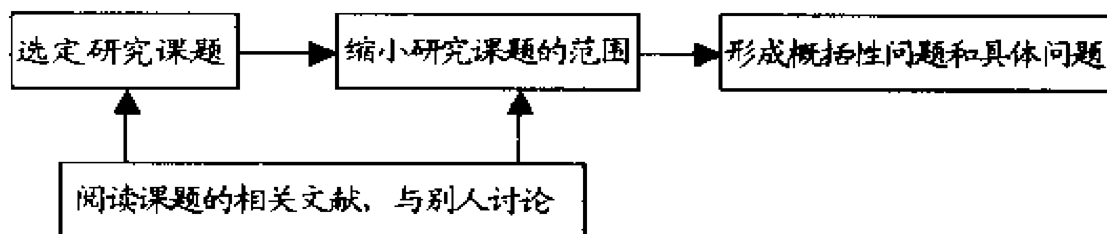


图 3.1：提出研究问题的程序

根据图 3.1，我们首先要找出所要研究课题的大致方向，然后不断缩小课题范围以确定研究焦点。无论是提出课题，还是寻找课题焦点，都需要广泛阅读相关的文献，适时地与导师或他人讨论。一旦确定了研究范围，研究者就需要提出概括性问题和具体问题。下面我们将逐一解释如何完成上述任务。

1. 选定研究课题

提出研究课题的第一步是在文献中寻找研究领域。你可能对第二语言教学或第二语言学习感兴趣，这两个领域是相关的，但两者有各自的

侧重点，研究的视角也不同。若以第二语言学习作为选题的起点，第二步缩小课题的任务就会非常繁重。如果选题的起点为第二语言写作，或更具体一点，是第二语言的议论文写作，缩小课题的任务就要比前面一个选题容易得多。很显然，选题的起点很重要。一开始选择的课题越具体，进入第二步（即缩小研究课题）就越轻松。选题时应该注意哪些方面呢？我们认为有 3 个方面值得注意。首先，个人兴趣应该放在首位。原因很简单，研究生学习并不是一项轻松的任务。你可能要在计算机里输入成千上万的问卷调查数据；可能要将几十个小时的录音录像资料转写成文字材料；可能在图书馆里呆上几天却找不到你要的材料。你对选题的兴趣才是帮助你走完这段艰难旅程的最大动力。

第二，选题不要受研究者本人情感的误导。譬如，有些学生天生胆怯，容易焦虑不安，害怕和别人交流。如果这样的学生研究焦虑程度和英语口语水平的关系，该课题的研究就很可能要受到个人情绪的影响。更为严重的是，她可能按照自己的感受去寻找研究答案（Ambady & Rosenthal，1992）。这样，收集数据时就可能出现偏差，碰到与自己感受不一致的建议也很难听得进。如此的研究存有明显的个人偏见。建构主义者认为，一切研究都带有主观性，要做到绝对客观既不可能，也不需要。需要强调的是，这里的情感误导指的是研究者的个人偏见，而不是一般意义上的主观性。

第三，应“避免选题过大和过难”（Rudestam & Newton，1992：10）。做论文，选课题，切忌贪大求全。要“小题大做”，而不要“大题小做”。许多研究生做研究刚起步，没有什么经验，但总想致力于一些标新立异的研究，希望在自己研究的领域内能取得突破性的进展。应该说，这些学生精神可嘉，愿望可以理解。但如果论文的目标定得过高，就会不切合实际，最终的命运只有两种：要么半途而废，要么质量低下。因此，不是目标定得越高越好，而是要选择切实可行的目标。做论文不能光凭热情，还需要脚踏实地的精神和实事求是的态度。即使是一篇优秀论文，研究者也少不了在自己期望的目标和实际能够达到的目标之间做出妥协，做出调整。但这并不是说我们鼓励学生去选择简单容易的课题，去撰写平庸低劣的论文。

总之，我们建议在确定论文题目时，选题要适度，切入点要准，要力所能及。在确信论文选题既不缺乏研究的原创性与新颖性，又不缺乏研究的广度和深度的前提下，所选的题目宜小不宜大。

2. 缩小研究课题的范围

一般说来，对研究者，特别是新手，缩小研究课题的范围是最棘手的问题。选择多大的题目才能做一篇像样的硕士论文或博士论文，这并没有十分明确的量化标准。有经验的导师往往在这方面都能做出判断并能给学生有益的指导和帮助。

缩小研究问题范围的时候，可以从“谁”、“什么”、“如何”等问题入手。假设你的选题是“英语词汇的习得”，这个题目显然太大。这时你可以问问自己，“学习者是谁?”他们是中学生还是大学生？如果是大学生，要区分是英语专业还是非英语专业；如果是英语专业，还要区分哪个年级。我们不妨选定英语专业大二学生。接着就是第二个问题：你感兴趣的是“什么”词汇？是接受性词汇还是表达性词汇？如果你感兴趣的是表达性词汇，那么就要确定是口语词汇还是写作词汇。我们假设选定的是口语词汇。剩下的问题就是“如何”开展研究。你是否做访谈？是否要求学生回答问卷？是否观察学生在课堂上的口语表现？是否打算测量他们表达性词汇量？是单独访问每个学生，还是在实验室里对他们逐一进行口语测试？要不断问自己这一类问题，只有这样，课题的范围才会变得越来越小、越来越明确。这个过程什么时候才能结束，取决于你的研究目的。

3. 阅读相关文献，与其他研究者讨论

在研究过程中，研究者需要不时阅读相关文献，与别人交流讨论，但不同阶段的目的不同，阅读与讨论的侧重点也有所不同。我们在图3.1中提到，阅读文献与交流讨论应该贯穿于选择研究课题和确定研究重点的全过程。学生进入论文选题阶段，大致有两种情况，一种情况是，没有明确的研究兴趣；另一种情况是，通过自己的观察或对自身经验的反思，已经形成了初步想法。无论是哪一种，都需要阅读相关文献和与别人交流讨论。

设想你进入了论文选题阶段，脑中没有什么明确的、感兴趣的课题。这时你可以从浏览相关的外语教学学术期刊入手，如 *Applied Linguistics*、*TESOL Quarterly*，*Language Learning* 等。浏览时不必每篇都读。先浏览文章的题目和摘要，碰到感兴趣的论文，再仔细阅读文章的最后一部分。通常作者在最后一部分提出一些需进一步研究的问题。通过阅

读这一部分，你或许会从中找到一些令你感兴趣的课题。另外，文章的研究成果也能引发灵感。例如，Huang（1984）的研究结果表明功能操练（即从事交际活动）可以预测中国学习者英语口语能力。你可能会问：功能操练是否也能预测中国学习者的阅读和写作水平呢？顺着这一思路继续思考下去，你可能就会形成自己的选题。如果你进入论文选题阶段时，已经形成了选题的初步想法，这时阅读相关文献的目的不是寻找灵感，而是论证自己的选题是否有研究价值。你可以浏览权威学术刊物在过去 10 年间发表的相关论文，阅读每篇论文的摘要。在阅读了各个时期的文献后，你就能确定自己所选择的课题是否值得研究。

阅读文献固然对选题、缩小选题范围很重要，但它不能代替与别人的“交流”。首先文献不能和学习者进行双向交流，同时文献在一定程度上具有滞后性。“交流”能弥补阅读文献的不足。我国传统的教育思想十分重视“交流”的作用。例如我国最早的一部教育文献《学记》中写道：“独学而无友，则孤陋寡闻”，意思是闭门苦读，没有学友的帮助，则知识就不容易得到增长。荀子在《劝学篇》中也写道：“学莫便乎近其人”，意思是学习的途径没有比接近良师益友更省事了。孔子说，“三人行，必有我师焉”（源自《论语》），俗话也说“听君一席话，胜读十年书”。常与别人交流，尤其是和导师交流，你能够集思广益，收到事半功倍的学习效果。我们发现，一些学生不会充分利用周围的人力资源，他们往往独自苦思冥想，很少与他人讨论，因此丧失了向别人学习的机会、丧失了获得别人帮助的机会。

4. 形成概括性问题和具体问题

概括性问题只起一个蓝本的作用，即只能为你的研究提供大致的方向。具体问题是从概括性问题中派生出来的，与研究程序中的具体细节直接关联，如研究对象的选择、资料的收集和资料的分析。

不要急于把研究问题过早地定下来，这样做会限制研究多元发展的可能性。在确定具体研究问题之前，最好花一些时间思考，列出研究发展的各种可能性。你通常会经历这样的阶段：你研究问题的数量不断增加，之后又不得不进行压缩。这属于正常现象，不必担心。在 Punch（1998）看来，缺少这样的经历只能说明你一开始就没有把问题考虑深、考虑透。

形成初步的具体问题后，需要仔细检查每个问题的重要性。无关紧

要的问题一定要下决心舍弃。一般原则是，与其肤浅地回答许多问题，还不如深入地回答几个问题。一篇硕士或博士论文一般不可能解决5个或更多的概括性问题 如果你的论文有5个以上的概括性问题，最好考虑减少到3至4个，这样重点突出，集中精力，论文容易做得有深度。

最后，研究问题应按照合理的次序组合和编排。换言之，问题中必须有比较明显的层次关系，如图3.2所示。定量研究的问题一般说来应在实证性研究开展之前确定。这些问题有时用假设的形式提出，当然，直接提出问题也无妨。定性研究问题在数据收集阶段之前通常是概括性的，随着研究的逐步展开，具体问题也就随之出现。

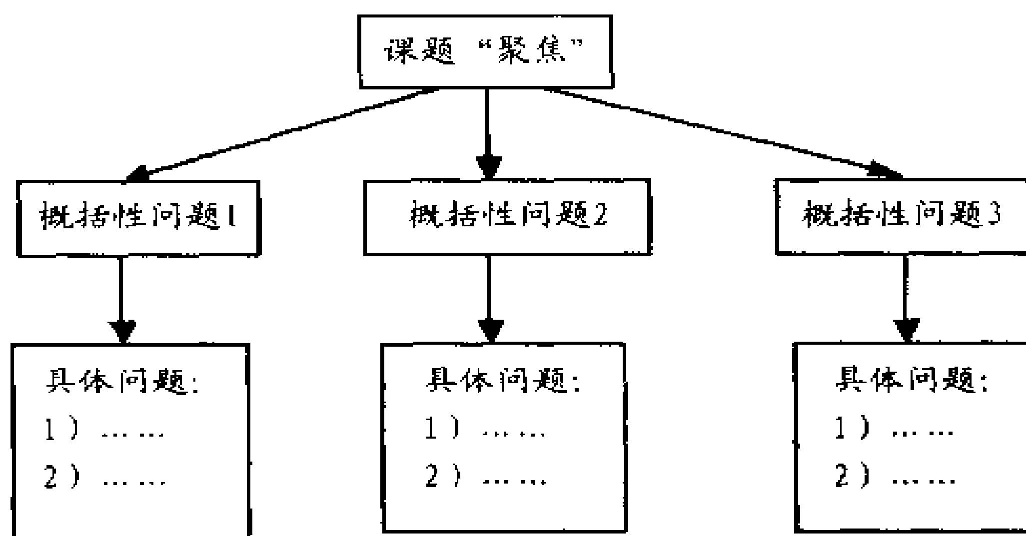


图3.2：概括性问题和具体问题的关系

下面的研究问题选自1998年南京大学一些研究生的毕业论文，我们稍作了改动：

例1（J.Wu，1998：3）

研究课题：第二语言议论文写作

课题焦点：大学生议论文英语写作研究：修辞知识和话语模式

概括性问题和具体问题：

1. 学生的修辞知识和话语模式之间存在什么关系？
   1）研究对象对英语议论文的话语结构的看法与英语本族语的话语模式有什么不同？
   2）研究对象撰写的英语议论文在修辞结构上与英语本族语的话

语模式有什么不同?

3) 研究对象的修辞知识和他们自己撰写的英语议论文习作是否相关?

2. 第二语言的水平是否会调整修辞知识和话语模式之间的关系?

1) 在观察英语议论文的修辞结构方面，不同语言水平的研究对象之间是否存在差异?

2) 他们在议论文的写作方面是否存在差异?

3) 如果研究对象按照其第二语言水平分组，那么他们的修辞知识和写作成绩之间有何关系?

例 2 (Xu, 1998: 14)

研究课题：EFL 写作

课题焦点：英语专业学生写作中的衔接

概括性问题和具体问题：

1. 第二语言的水平在多大程度上影响 EFL 写作中的衔接模式?

(a) EFL 写作中的衔接模式是否因学生二语水平的不同而变化?

(b) 如果是这样，高水平学习者与低水平学习者在衔接模式上有何差异?

2. 第二语言的水平对衔接错误模式影响程度有多大?

(a) EFL 写作中的衔接错误模式是否因学生二语水平的不同而变化?

(b) 如果是这样，高水平学习者与低水平学习者在衔接错误模式上有何差异?

3. 衔接错误模式在多大程度上影响写作质量?

(a) 不同年级之间的衔接模式与衔接错误模式的差异是否会引起写作质量的变化?

(b) 如果有变化，质量好的作文与质量差的作文之间在衔接模式与衔接错误模式上有何差异?

(c) 哪些衔接手段更能预测写作质量?

例 3 (Zhu, 1998: 11－12)

研究课题：第二语言的口语水平

课题焦点：不同任务和计划时间的长短对二语口语水平的影响

概括性问题和具体问题：

1. 对任务的熟悉程度和二语口语水平之间有何关系?

假设 1:

学生对最熟悉的任务所产出的语言准确程度可能最高，而对最不熟悉、最难的任务，学生口语的准确程度可能最低。

假设 2:

研究对象对最不熟悉、最难的任务所产出的语言，其复杂程度可能最高；相反，研究对象对最熟悉的任务所产出的语言，其复杂程度可能最低。

假设 3:

研究对象对最熟悉的任务所产出的语言流利程度可能最高，而对最不熟悉的任务，研究对象产出的语言流利程度可能最低。

2. 计划时间的长短和第二语言的口语水平之间有何关系？

假设 1：计划时间越长，语言的准确程度越高。

假设 2：计划时间越长，语言的复杂程度越高。

假设 3：计划时间越长，语言的流利程度越高。

第二节 选题、定题过程中的困难

前面提到，选题、定题通常是研究新手碰到的第一道难关。在多年研究生论文指导实践中，我们发现学生常常碰到的困难有：1）没有研究想法；2）选题过大、过泛；3）选题没有研究价值；4）研究问题像购物单，缺乏内在的逻辑性。

1. 没有研究想法

有些研究生在完成了所有学位课程后还没有具体的研究想法。等到需要拿出研究计划时，他们不知道准备做哪一方面的研究。他们以为导师可以为他们选择研究课题。这种想法是错误的，也不切合实际，原因很简单：导师，顾名思义，是负责指导和监督你的研究，而不是替你做研究。况且，学会独立选、找研究课题是研究生学习的一部分。培养这种能力，将使你终生受益。否则，即便完成了学位论文，你的科研生涯也将由此结束。有一些导师缺乏指导经验，出于好意，为学生提供或推荐课题，我们认为这样做弊大于利。首先，学生在选题、定题方面没有得到适当的培训，研究能力的培养就不全面。此外，这样的做法至少还

有两个弊端。首先，一篇硕士或博士论文通常需要几年的时间才能完成。根据我们的经验，从事论文研究的过程中一般会有一段时期感到枯燥、厌烦，甚至对研究的课题逐渐丧失兴趣。如果他们的课题是别人提供的，这种消极情绪会更加强烈。有些学生甚至感到后悔，后悔当初不该选择导师提供的课题。其次，研究过程中难免有困难。如果课题是导师提供的，一碰到困难，他们就会找导师，因为他们认为这是导师的课题。因此，我们不赞成导师提供课题。研究生训练是一个完整的周期，选题、定题是整个周期中不可或缺的环节，这一点，我们特别要强调。

怎样才能选定一个有意义的课题呢？广泛阅读文献是研究灵感的重要源泉。除此之外，还有两个重要的来源：个人经验和与别人交流。

1.1 个人经验

个人学习第二语言的经历本身就是寻找科研题目的重要源泉之一。然而，没有强烈的好奇心和很高的悟性，就不能从自身的经历中获得灵感。但是好奇心和悟性不是与生俱来的，这需要后天的培养。因此，我们建议，无论什么时候对语言学习和教学有疑问或想法，都要将它们记录在专门的笔记本上。这些想法就像闪烁的火花，稍纵即逝。抓住这些火花，它们就可能引发出一个有价值的科研项目。比如说，即使成功的英语学习者在口语中也常常混淆人称代词 he 和 she，误用冠词也不足为怪。这时，你可以问这样的问题：中国学习者学习英语最难掌握的语言形式是什么？在掌握这些语言形式时，中国学习者之间有何差异？只要你勤于思考、多问几个为什么，那么从你自己的第二语言学习经历中就可以发现无数值得研究的课题。

1.2 参加学术会议，与别人交谈

研究课题的第二个来源是参加学术会议、与其他研究者交谈。已发表的论文所报告的通常是研究者两三年前所作的研究，但是学术会议上报告的通常是最近的研究成果或正在进行的研究项目。参加学术会议，不仅能够获得有关领域研究的最新动态，还可以结识一批知名学者。每次发言后，会议一般都要安排大会提问，有时还安排小组讨论，这时你可以向发言的学者直接提问，也可以会后向发言者当面讨教。这种交流能激发你研究的灵感，帮助你找到解决难题的方法。你要不失时机地参加相关的学术会议，也要认真听取本校请来的校外专家作的讲座。重要的是，要抓住机会和他们交谈，与他们交朋友。

2. 选题过大、过泛

研究新手一般经验不足，选题往往过大、过泛，缺乏研究重点。如果这是在研究的初始阶段，就算不上一个问题，因为好选题的形成需要有个过程。但如果研究者根本不知道自己选题中存在的问题，那就麻烦了。比如，有一个学生想研究情感因素对非英语专业研究生英语学习的影响。这个问题似乎是一个有趣的话题，以往的研究主要集中在认知因素上，只有少数人研究了一两个情感因素。从这个意义上讲，这个课题值得研究。困难的是他没有意识到这个题目太大，需要进一步缩小范围。

他认为自己在缩小题目范围上，已经做出了努力。比如，他选择了情感因素，而不是与学习者相关的所有因素；研究对象是非英语专业研究生，而不是所有的在校大学生。可他没有想到，情感因素很多，一项研究很难面面俱到，他也没有意识到，几乎没有什么有效的工具来测量情感因素。最后，在导师的帮助下，他缩小了研究课题，只研究其中一个情感因素，即学习动机。

3. 问题没有研究价值

有些学生提出的问题没有研究价值，这些问题既没有实际意义，又没有理论价值。你会说，这怎么可能呢？的确，没有哪个人愿意浪费大量的时间和精力去研究没有价值的问题。然而，研究新手很难看出什么样的问题是鸡毛蒜皮的问题。例如，有一个研究生想观察英语专业一年级学生和三年级学生在写摘要方面有什么差异。这个概括性的问题听上去很有趣，似乎不可能没有研究价值。但一看具体问题，就知道为什么我们认为他的研究没有价值。他的具体问题包括：1）这两组学生在写摘要方面的总体成绩是否存在差异？2）这两组学生的摘要里所包含的要点数是否存在差异？3）这两组学生的摘要里所包含的语法错误的总数是否存在差异？所有这些问题提得似乎合情合理，每个问题都有明确的侧重点。他花费了大量时间评定摘要的成绩、统计摘要的要点数和语法错误。然而，不用他提供实证数据，我们就能回答这些问题：三年级学生的分数应该比一年级高，他们摘要里的要点数肯定比一年级学生多，语法错误也肯定比一年级少。这些结果对改善教学和学习有何启发呢？对理论建构有何贡献呢？答案是否定的。由此可见，这个学生提出

的问题没有研究价值。从上述例子我们可以看出，即使研究的过程非常复杂，也有实证数据支撑，研究问题也有可能没有价值。一个人闭门造车最容易犯这样的毛病。克服这一毛病最直接的方法是与别人讨论，广泛征求意见。

4. 研究问题像购物清单，缺乏内在逻辑性

有些研究生常常会列出一组问题，这些问题既没有概括性问题与具体问题的区别，又没有明显的逻辑关系。我们把这样的做法比喻为“列购物清单”。出现这样问题的主要原因是研究者没有进行认真、充分的思考。很显然，这样的问题需进一步修改、删减和整合，才能作为论文最终确定的问题。

陷入了这样的困境，该怎么办？我们建议，先列出概括性问题，然后找出各问题间的联系。这样做的前提是，你对论文选题具有比较全面、透彻的了解。问题之间的逻辑联系非常重要，它能使论文构成一个有机的整体。没有这样的关系，论文就会有拼凑之感。确立逻辑关系，首先要考虑概括性问题之间的联系。大问题弄清楚了，确立具体问题之间的关系就相对容易。

请看下面的例子。第一组问题是研究初始阶段形成的，第二组是经过仔细修改和润色之后重新组合而成的。

第一组问题是研究课题的雏形：

1）母语在第二语言的写作过程中是否有作用？如果有，母语在第二语言写作中有多大的作用？
2）在第二语言写作过程中，母语何时介入？
3）母语在第二语言写作过程中具有什么样的功能？
4）在第二语言写作过程中，为什么学习者使用母语？
5）在第二语言写作中使用母语是否随着学习者第二语言水平的发展而变化？
6）在第二语言写作中使用母语是否随着写作任务的类型不同而变化？
7）在第二语言写作过程中使用母语和写作的质量之间是否有相关性？
8）第二语言学习者对在第二语言写作中使用母语有何看法？

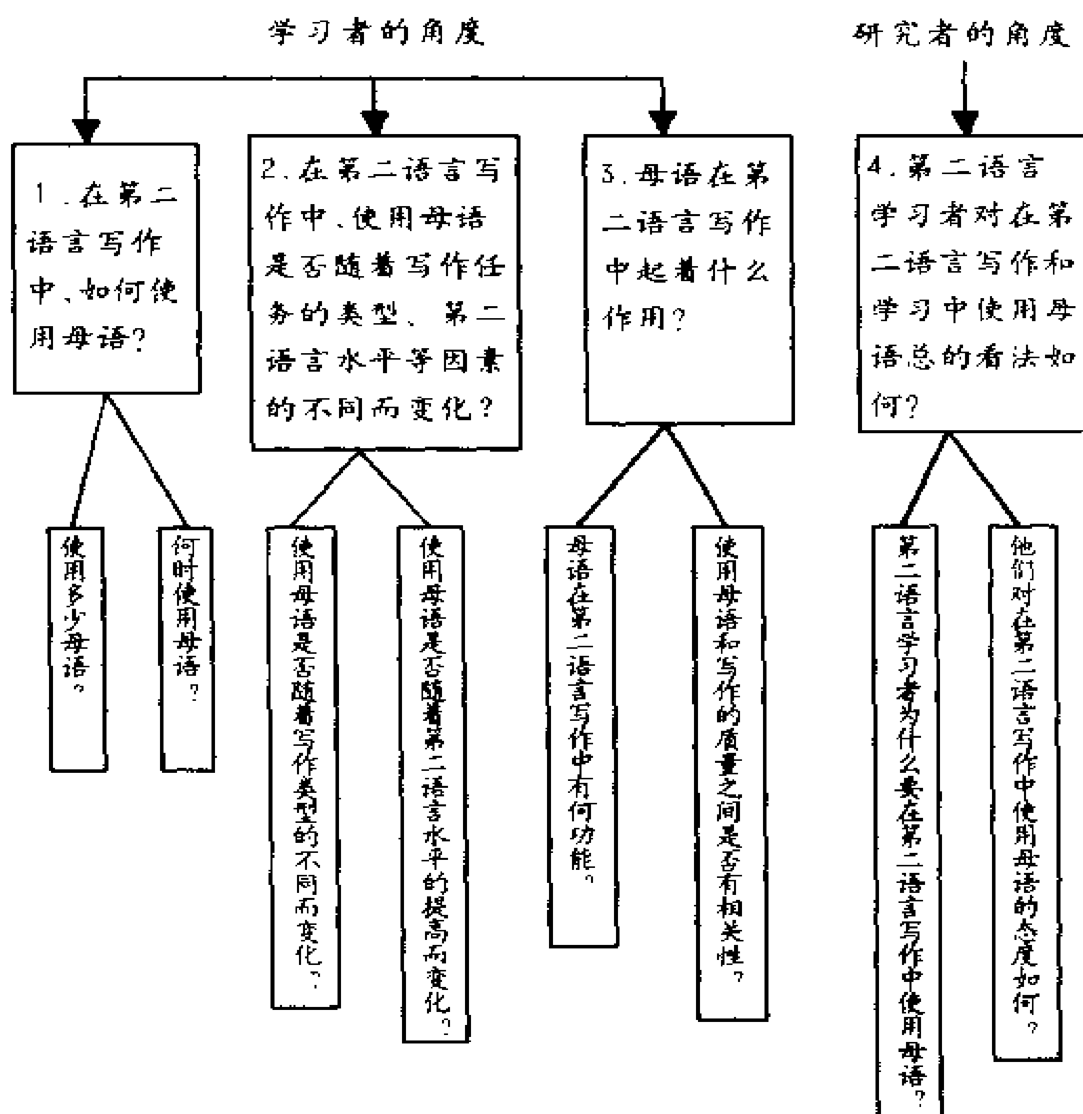


图 3.3：第二组问题图示说明

第二组问题是经过修改调整后确定的问题：

1）在第二语言写作过程中，英语专业学生如何使用母语？
　（1）母语在第二语言写作中有多大作用？
　（2）在第二语言写作过程中何时使用母语？

2）使用母语是否随着写作任务的类型、第二语言水平等因素的不同而变化？
　（1）在第二语言写作过程中使用母语是否随着写作任务的类型不同而变化？
　（2）在第二语言写作过程中使用母语是否随着学习者第二语言水平的发展而变化？

3）母语在第二语言写作中起着什么作用？
　（1）母语在第二语言写作过程中具有什么样的功能？
　（2）在第二语言写作过程中使用母语和作文的质量之间是否存在相关性？
4）第二语言学习者对在第二语言写作和学习中使用母语总的看法如何？
　（1）在第二语言写作过程中，为什么学习者使用母语？
　（2）第二语言学习者对在第二语言写作中使用母语的态度如何？

研究者在第一组问题里提出了8个与使用母语相关的问题，这8个问题的排列让读者很难看出它们之间的内在联系。而经过重新组合的第二组问题，分成了4个大问题，每个大问题包含两个小问题。第一个问题提出研究第二语言写作中使用母语的总体情况；第二个问题旨在发现母语的使用受写作话题、第二语言水平等因素影响的程度；第三个问题旨在说明母语在第二语言写作中的具体功能；最后一个问题目的是调查第二语言学习者对在第二语言写作中使用母语的态度。由此，这4个大问题之间的内在关系就一目了然了。

小结

选择研究课题是第一步。第二步要缩小课题的范围，找到研究课题的“聚焦点”。缩小课题范围的最好的方法是反复地问以“谁”，“什么”，“如何”开头的这类问题。在寻找研究课题和确立研究焦点的过程中，一定要阅读相关文献，与其他研究者交谈讨论。确定了研究的焦点后，需要把研究课题转变为几个概括性问题和具体问题。在选题、定题的过程中，研究新手常碰到的困难有：没有研究想法；选题太大；选题没有研究价值；研究问题杂乱无章，像一份购物单。克服这些困难的办法多种多样。我们建议多反思自己学习外语的经验，多参加学术会议，多与别人交流讨论，以获得专家学者的及时帮助与指导。

讨论题

1．选题、定题的一般程序是什么？
2．选题、定题时，首先应该考虑什么？
3．用什么方法来缩小研究课题的范围和确定研究的焦点？
4．阅读文献在确定研究问题方面发挥什么样的作用？

5. 为什么在提出概括性问题的同时还需要提出具体问题?
6. 请试着缩小下面的研究课题的范围:
   (a) 英语冠词的学习
   (b) 中国学习者掌握情态动词的方法
   (c) 英语时态系统的教学
   (d) 第二语言议论文的教学
   (e) 母语的使用和第二语言学习之间的关系
7. 对下面所给的两组问题进行分类，所作分类要求显示出概括性问题和具体问题之间的逻辑关系，如果有必要，你可以增加或删除个别问题(注：下面的问题选自学生的论文，略有改动)。
   1) 中国英语专业学生冒险观念和行为的研究
      (1) 中国学生对在英语学习中的“冒险”持什么态度?
      (2) 他们冒险行为的平均频率是多少?
      (3) 他们的冒险观念和冒险行为之间有何差异?
      (4) 高、中、低程度的冒险者对冒险的看法分别是什么?
      (5) 高、中、低程度的冒险者在英语学习中分别采取什么行为?
      (6) 高、中、低程度的冒险者在观念和行为上有何差异?
      (7) 冒险和第二语言的学习结果之间是否有相关性?
      (Wang, 1999)
   2) 大学英语课堂教学中纠错的研究
      (1) 精读课教师如何纠正学生的错误?
      (2) 英语专业教师和非英语专业教师在纠错方面是否存在差异?
      (3) 学生对纠错持什么态度?
      (4) 在纠错态度和纠错方式的偏好上，英语专业学生和非英语专业学生之间是否存在差异?
      (5) 就纠错方式而言，教师的行为在多大程度上与学生的期望相符合?
      (Hu, 1999)

第四章　阅读文献

我们在上一章提到了选择选题时可以在文献阅读过程中获得灵感，此外，文献阅读还可以帮助确认选题有无研究价值。这一章将介绍阅读文献的其他用途。

第一节　文献来源

文献资料大体上可分为两类：一类是应用语言学领域的文献，另一类是相关领域的文献，如语言学、社会语言学、心理语言学、教育学、心理学和社会学等。阅读文献的范围要广，内容要全面，既要阅读应用语言学领域的文献，又要阅读相关领域的文献。我们应该知道，应用语言学研究的发展常常得益于相关领域的理论与研究方法。就一个研究项目而言，究竟要阅读多少文献为好？这个问题将在“确定阅读文献的范围”一节里讨论。

文献来源多种多样，如索引、会议论文、学术论文、著作。下面我们将逐一介绍。

1. 索引

有些索引每个季度出版一次，有些索引每月出版一次，也有些每一年或半年出版一次。这些索引是文献的最好来源。表 4.1 列出了应用语言学领域最常用的索引。

索引为读者提供各种研究课题的参考书目，这些书目根据作者的姓名和研究的主题编排，十分便于查找。参考书目中包括多种有用信息，如出版年代、出版地、出版社名称、索取地点和方式、出版形式（论文、文章或报告）等。大多数索引还包括有关资料的摘要或小结。

索引名称	出版刊次频率	出版地点	提供的信息
Language Teaching	每季	Cambridge: Cambridge University Press	学术刊物的文章、著作和报告
Resources in Education	每月	Washington, DC: Government Printing Office	文章、报告摘要、会议论文及其他研究报告
Current Index to Journals in Education	每月	Phoenix, AZ: Oryx Press	约 780 种学术期刊中与教育相关的文章

表 4.1：索引举例

2. 未发表的论文

未发表的论文* 指会议论文或在同行中传阅、供讨论的手稿。这类论文是我们了解正在进行的或最新完成的科研项目的重要来源。在这类论文里，除了标题和摘要之外，通常还可以了解到论文作者的姓名，通讯地址和电子信箱地址。我们如果需要帮助，可以很容易地与作者取得联系。这类论文还可以提供最新的参考文献，而已经发表的论文常常在编辑部就已经搁置了一两年。

3. 学术期刊

学术期刊包含定期发表的论文，不同的期刊面对不同的读者群。譬如在第二语言习得的学术圈内，有些刊物的读者主要是研究者；有些刊物的读者是教师和管理者；还有些刊物，两种读者的需要都考虑。下面列举的刊物是应用语言学领域里的核心期刊：

Applied Psycholinguistics

Applied Linguistics

---

* 未发表的论文和发表的论文一样也应该列在参考书目里。与已发表的论文不同的是，你无需注明出版社或刊物名称，只需标出"手稿"名称，在何时何地传播过（如课程讲义，哪种课程；会议论文，哪一次会议；硕士论文，哪一所大学），同时论文应给出网址。否则，将视为剽窃。

Annual Review of Applied Linguistics
Asian Journal of English Language Teaching
Canadian Modern Language Review
ELT Journal
English Teaching Forum
English for Specific Purposes
Foreign Language Annuals
Interlanguage Studies Bulletin
Language Testing
Language Learning
Modern Language Journal
Studies in Second Language Acquisition
System
TESOL Quarterly
TESOL Journal

在上面列出的期刊中，*Applied Linguistics* 和这类的刊物主要面向研究者；*English Teaching Forum* 和 *ELT Journal* 主要面向实践者；而 *System* 和 *TESOL Quarterly* 既面向研究者，又面向实践者。

4. 著作

提供参考文献的著作有 4 种：1）教材，2）语言学词典和百科全书，3）学术专著，4）论文集。

教材的优点是：对相关领域能作全面的介绍，对一些概念的阐述比论文更具有系统性。教材可以就论文中常常出现的某些主题展开讨论和评述，而这些正是学生所缺乏的。由于篇幅所限，论文一般没有背景介绍。

对于一些主要概念的定义和这些概念的研究起源，查考语言学百科全书和语言学词典这类工具书特别有效。在查寻近期的文献时，你可能会获得最新的争议焦点或相关定义，但它们不是最初来源。对你最有帮助的语言学百科全书和词典有 David Crystal 主编的 *The Cambridge Encyclopedia of Language*，Tom McArthur 的 *Oxford Companion to the English Language*，Matthews 的 *The Concise Oxford Dictionary of Linguistics*，以及 Richards, Platt & Platt 主编的 *Longman Dictionary of Language Teaching and Applied Linguistics* 等。

学术专著是作者自己就某一个课题发表的理论或观点，对某领域或某种理论的发展有时具有很高的权威性，如 Krashen 的著作《第二语言》(*Language Two*)。

论文集可以是一次学术会议的论文汇编，也可能是编辑组稿的结果，也有可能是出于合作需要就某个跨学科主题进行研究的成果。论文集汇集了众家思想，是许多专家对某个中心主题的思考和见解。论文集主题明确，视角独特，对我们的研究有事半功倍的启发作用。与索引、会议论文和学术刊物相比，著作作为参考的来源有其自身的弱点。比如，著作里的参考文献不是最新的，毕竟，编写一本书所花的时间要比撰写一篇论文要长得多。再者，著作带给我们丰富的背景知识，而期刊带给我们的是最新的研究成果。

第二节　查询、阅读、综述文献的一般程序

研究新手走进图书馆的时候常常感到不知所措，对周围大量的资料无从下手。在这里，我们向大家推荐一套行之有效的查询、阅读、综述文献的程序（见图 4.1）。需要说明的是，这不是惟一的方法。

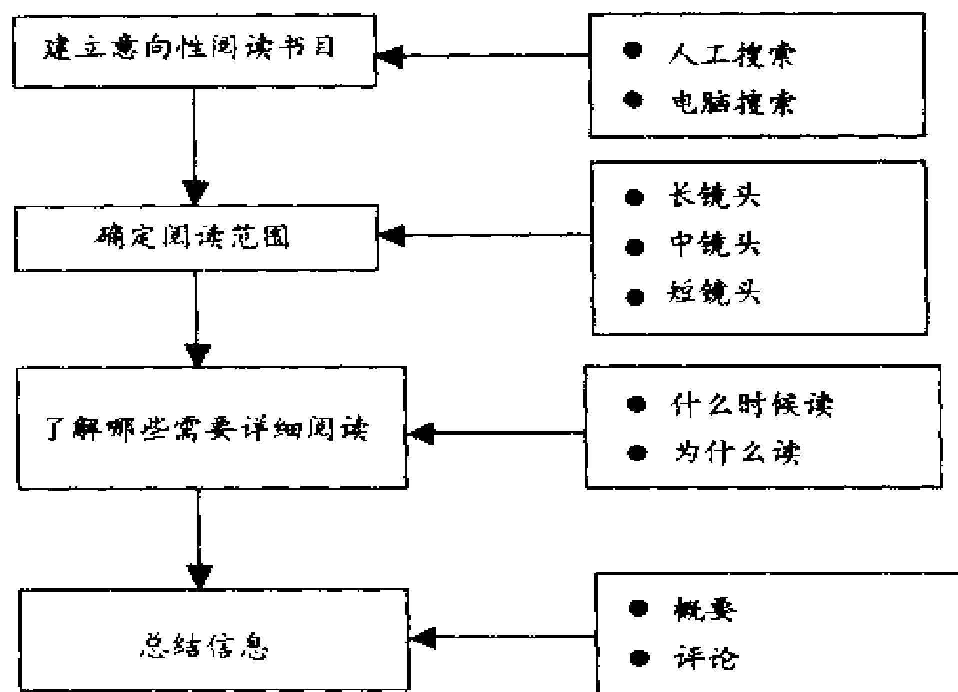


图 4.1：查询、阅读、综述文献的一般程序

1. 建立意向性阅读书目

1.1　什么是意向性阅读书目?

所谓意向性阅读书目指的是研究者列出的可能需要阅读的参考文献清单。建立这样的参考书目有两个目的：第一，它可以作为“阅读文献”的指南；第二，它可以充当一种资源库，作为撰写硕士或博士论文“引用文献”的依据。一般来说，无论是硕士论文，还是博士论文，引用文献都是在不断修改中完成的，所以论文中的引用文献也要随之变化。今天所引用的文章，明天可能会删除，后天又可能补上。每次修改论文之后都要变化一次“引用文献”，这样做会事倍功半。最省事的办法是，论文修改完毕后，从你建立的意向性阅读书目中选择那些被引用的论文或著作。

意向性参考书目中的文献应按首字母排序，每个项目应包括作者的姓名、作品名称、出版地点、出版社名称等。如果是期刊论文，参考书目应包括期刊名称、卷号和页码范围；如果是著作，参考书目应包含出版社、出版地点等方面的信息。参考书目要认真校对，保证每一项准确无误。

1.2　人工搜索

人工搜索，顾名思义，指研究者自己动手确立所需的研究资料，而不借助于电脑。根据我自己的研究经验，人工检索使用滚雪球的方法比较有效，也就是说，从少数几个参考文献人手，每个参考文献会引出更多的参考文献，这个过程几经重复，参考书目的数量就会成倍增加。具体地说，第一步，从一篇与你研究课题最相关论文中的参考文献着手，剔除参考文献中那些无关的书目。通过筛选，参考文献中可能会剩下 10 到 15 个书目，这些书目中也许在本校图书馆里能找到 5 到 10 个。到了第二步，阅读这些论文的摘要，判断哪些论文与你的选题相关。假设只有 5 篇论文对你有价值，那么，这 5 篇论文保留下来，从书目中去除其他 5 篇。然后，按照第一步的程序，从这 5 篇论文后面的参考文献中筛选出与你的课题明显相关的参考书目，去掉那些无关的书目。如此循环往复，参考书目的数量就会呈指数式增长，最终形成一份意向性阅读书目。

在这个阶段，有一点必须牢记，你只需阅读每篇论文的摘要，没有

必要通读论文全文。阅读全文会使你陷入论文的细节，这样做既浪费时间，效果又差，而通过比较每篇论文的摘要，你会更全面地看到文献与你课题的相关性。

1.3　计算机搜索

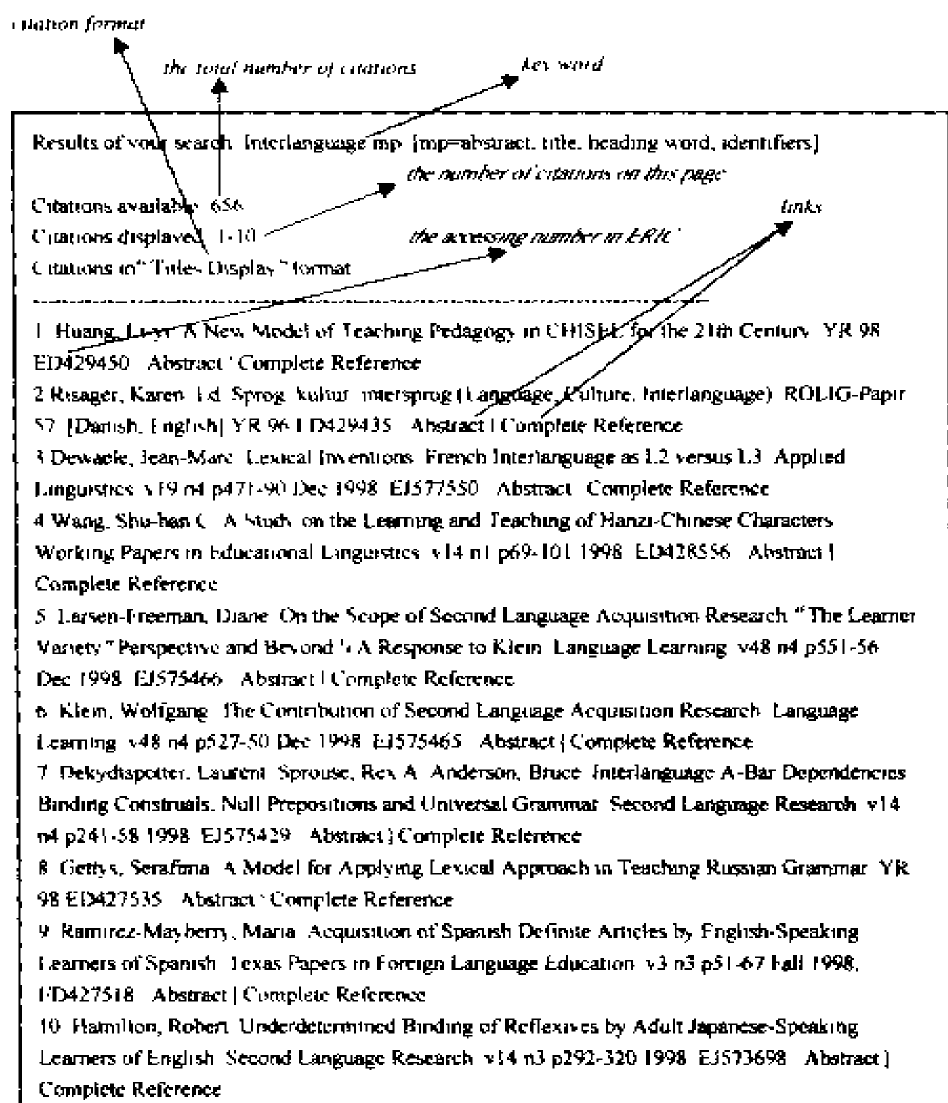
利用计算机搜索文献是一种简单、快捷的方法，意向性阅读书目的建立可以依据计算机数据库提供的信息。近年来，国内很多大学图书馆购买了网络数据库，如 CNKI 数据库中的“中国学术期刊全文数据库”，Elsevier，EBSCO（全文期刊数据库），Kluwer Online，ProQuest Digital Dissertation（PQDD）以及 Cambridge Scientific Abstracts（CSA）中的 Linguistics and Language Behavior Abstracts（LLBA）等。这些数据库是利用计算机检索文献的很便利的资料源，但这些数据库往往不能自动生成参考书目。目前最有影响的第二语言习得数据库是由“美国教育部教育资源信息中心”（英文字母缩写是 ERIC）建立的。这个数据库包括700,000 项参考书目和论文摘要，索引每年覆盖 800 种学术刊物和13,000 种文件（包括所选著作）。

运用 ERIC 时，只要输入一个关键词，数据库中就会生成一份相关的参考书目。比如，输入关键词 interlanguage，在电脑屏幕上马上就会出现如图 4.2 所示的 ERIC 数据库内容。搜索结果包括两个部分的内容。第一部分报告的是搜索的一般信息，以 Title Display（标题显示）格式出现的索引共计 656 条，第一页只出现了 10 条。第二部分报告的是索引清单，所列的每个项目包括作者、标题、出版地、何种语言撰写、进入 ERIC 的号码等。在每个索引项最后，有两个链接：Abstract（摘要）和 Complete Reference（完整的参考文献），只要点击鼠标就能进入。

与人工搜索不同，计算机搜索能在几秒钟内生成一份参考书目，但这种书目不是我们所说的意向性阅读书目。计算机生成的参考书目至少需要经过 3 次筛选才能删除那些无关的书目。第一次筛选很简单，删除那些明显没有价值的索引项目；第二次筛选的方法是通过点击链接 Abstract，阅读相关论文的摘要；最后，搜索计算机图书系统，检查所需资料能否获取。经过 3 次筛选留存下来的书目即可以进入意向性阅读书目。

计算机搜索只需要坐在计算机前，点击鼠标，从这个意义上讲，计算机搜索省时省力，而且资料库中的资源丰富，内容新颖。因此，通过计算机搜索建立起来的意向性阅读书目要比通过人工搜索产生的更全

面。但是，ERIC 以及其他基于计算机的资源也存在许多问题，如数量太多，质量没有保证，需要耗费大量时间剔除那些无关的文献。



Results of your search: Interlanguage.mp. [mp=abstract, title, heading word, identifiers]

Citations available: 656
Citations displayed: 1-10
Citations in "Titles Display" format

1 Huang, Liyu. A New Model of Teaching Pedagogy in CHISEL for the 21th Century. YR 98 ED429450 Abstract | Complete Reference
2 Risager, Karen. Ed. Sprog, kultur, intersprog (Language, Culture, Interlanguage). ROLIG-Papir 57. [Danish, English] YR 96 ED429435 Abstract | Complete Reference
3 Dewaele, Jean-Marc. Lexical Inventions: French Interlanguage as L2 versus L3. Applied Linguistics. v19 n4 p471-90 Dec 1998. EJ577550 Abstract | Complete Reference
4 Wang, Shu-han C. A Study on the Learning and Teaching of Hanzi-Chinese Characters. Working Papers in Educational Linguistics. v14 n1 p69-101 1998. ED428556 Abstract | Complete Reference
5 Larsen-Freeman, Diane. On the Scope of Second Language Acquisition Research: "The Learner Variety" Perspective and Beyond: A Response to Klein. Language Learning. v48 n4 p551-56 Dec 1998. EJ575466 Abstract | Complete Reference
6 Klein, Wolfgang. The Contribution of Second Language Acquisition Research. Language Learning. v48 n4 p527-50 Dec 1998. EJ575465 Abstract | Complete Reference
7 Dekydtspotter, Laurent; Sprouse, Rex A.; Anderson, Bruce. Interlanguage A-Bar Dependencies: Binding Construals, Null Prepositions and Universal Grammar. Second Language Research. v14 n4 p241-58 1998. EJ575429 Abstract | Complete Reference
8 Gettys, Serafima. A Model for Applying Lexical Approach in Teaching Russian Grammar. YR 98 ED427535 Abstract | Complete Reference
9 Ramirez-Mayberry, Maria. Acquisition of Spanish Definite Articles by English-Speaking Learners of Spanish. Texas Papers in Foreign Language Education. v3 n3 p51-67 Fall 1998. ED427518 Abstract | Complete Reference
10 Hamilton, Robert. Underdetermined Binding of Reflexives by Adult Japanese-Speaking Learners of English. Second Language Research. v14 n3 p292-320 1998. EJ573698 Abstract | Complete Reference

(Note: The italicized words are explanatory notes made by the author.)

图 4.2：ERIC 生成结果范例

第三节 确定文献阅读范围

完成了意向性阅读书目，下一步就是确定文献阅读的范围，也就是说，划出阅读文献的界限。如果阅读文献的范围太宽，会使人望而却步，不知所措，最后迷失方向；如果面太窄，一些重要文献就会遗漏。

因此，划定阅读文献的界限必不可少。这里，我们将就如何划出这样的界限提出建议。

1. 远镜头、中镜头和特写镜头

所谓划出阅读文献界限就是决定阅读文献的范围及阅读的详略程度。我们知道，画家在画风景时，动笔时不会一开始就注重细节，而是先勾勒出草图，规划出草图中的前景、中景和背景。确定阅读文献的范围也是如此。在界定阅读范围之前，我们首先要区分 3 种不同的文献综述：背景式综述，前景式综述和中景式综述。

为了便于叙述，我们不妨借用电影拍摄中的术语来讨论上述 3 类综述：远镜头、中镜头和特写镜头（Rudestam & Newton，1992）。在摄影中，这 3 种不同镜头的区别在于摄影机和物体之间的距离不同。在文献综述中，3 种文献综述的区别主要在于描述细节的程度不同。

远镜头或背景式综述通常给研究项目提供一般的研究框架或给某些变量提供理论上的界定。特写镜头或前景式综述注重某些研究项目的细节，它们是研究设计的基础。中镜头式的综述主要概述与本课题相关的实证性研究。不过，这 3 类综述的差别也是相对的，对它们的理解要根据不同情况。换言之，没有具体的情形，很难说清楚什么是远镜头，什么是中镜头，什么是特写镜头。

我们用下面 3 个例子来说明上述 3 种综述的差异。

例 1

研究课题：	英语专业二年级学生的冒险观念与第二语言运用之间的关系
远镜头：	教育理论方面的资料，论述冒险行为与总体成绩之间的关系
中镜头：	第二语言习得研究领域里的资料，从理论和实证的角度讨论冒险行为与第二语言运用之间的关系
特写镜头：	国内外有关该课题的相关实证性研究

例 2

研究课题：	学习者可控因素对英语专业二年级学生成绩的影响
远镜头：	有关影响学习结果因素的教育理论
中镜头：	第二语言习得研究领域里有关该课题的理论模式；

有关在中国环境下英语教学与学习的资料

特写镜头： 国内外有关该课题的相关实证性研究

例 3

研究课题： 在第二语言写作过程中，母语有何作用

远镜头： 有关在第二语言学习中使用母语的理论阐述

中镜头： 与该课题相关的实证研究

特写镜头： 有关在第二语言写作中使用母语的实证性研究

上述 3 个例子清楚地告诉我们，“远镜头”综述通常与理论阐述有关；“特写镜头”只关心与你自己课题设计直接相关的实证性研究；“中镜头”综述可以是理论方面的，也可以是实证方面的。它们可以排列在两个连续体上：1）从抽象到具体；2）从具体到抽象。

如果我们能弄清楚某个研究项目所需要的上述 3 种不同综述，我们就比较容易确定阅读的范围了。

2. 确定需要详细阅读的内容

在对所选材料进行阅读之前，首先要弄清楚阅读重点该放在文献的哪一方面。对此我们没有直接的答案，因为阅读的内容取决于什么时候读，该读什么（见图 4.1）。

在研究的不同阶段阅读文献的目的有所不同。研究的初始阶段，阅读大量的研究文献是为了寻找研究领域，或者是为了确定自己的课题有无研究价值。换言之，这种文献阅读是为了选择课题和确立这个课题研究的必要性。一旦找到了研究课题，阅读文献的目的就发生了变化。这时的文献阅读为的是形成合适的研究问题和有效设计。到了论文的最后阶段，阅读文献是为了讨论研究结果的意义。如前所述，阅读文献贯穿于整个研究过程。通常，研究者每月至少去一次图书馆，阅读最新发表的论文，更新自己的意向性阅读书目。

第四节 整理与综合文献

有了充足的资料，下一步要做的就是阅读这些资料。我们建议使用逆向阅读法，即从最新的材料开始读起，然后慢慢往回读。那么，在阅

读过程中和阅读之后我们该做些什么呢？常见的做法是整理和综合阅读的内容，把它们记录在卡片上。最好准备3套卡片分别用于：记录摘要、记录评论、记录可能的用途。这么做看上去浪费时间，但这样的归档处理方法对日后的回忆和索取非常有用。在你进入论文写作阶段时，它的优势自然会显示出来。另外，这三套卡片必须可以交叉参照，便于你找到所需资料。

摘要卡片

摘要应该包含论文中最重要的信息。摘录摘要的卡片通常以字母顺序排列，也可以根据与研究课题相关的不同小标题排列。每个摘要集中在一张卡片上，或者在自己的电脑里建立一个数据库。每个摘要应该包括下面的信息：

1. 提出该研究课题的理由；
2. 该研究隐含的假设；
3. 收集数据的程序；
4. 该研究中的研究对象和测量工具；
5. 数据分析程序；
6. 该研究的主要发现；
7. 该研究对现有文献所作出的贡献。

大多数情况下，论文本身含有摘要，但如果摘要不能满足我们的要求，就得自己修改。有些论文没有摘要，就必须动笔写。为了节省时间，这种摘要不需要写成连贯的概要，只需用笔记的形式即可。换句话说，摘要里可以使用一些不完整的句子，摘录信息的多少应该以能否用于日后撰写文献综述为准，这是最基本的要求。

撰写摘要不是我们想像的那么难，但很耗费时间。有些研究生往往开始的时候还算认真，但是渐渐地失去了耐心，笔头也变懒了，他们觉得自己年轻，记忆力强，没有必要花那么多的时间去写摘要。然而，他们很快就会发现，他们的记忆并不是那么可靠，最终还是得回过头来重新阅读一大堆论文。要避免这样的麻烦，最好要养成良好的习惯，读完每篇论文，马上写出一份摘要。

评论卡片

对一项研究的评价包括该研究的优点和缺点，特别是论文中所存在的问题。我们建议，不要等到读完论文再动笔记下你的发现，这是因为这些发现常常是昙花一现，稍纵即逝，在这种情况下，马上把你的想法

写下来，事后再将它们组织到卡片上。

摘要是描述性的，而评论是评价性的，所以写评论不如写摘要那么容易，因为写评论首先要用批评的眼光仔细阅读材料。研究对象的选择，测量工具的设计，数据的收集与分析，对研究结果的解释等，这些都有可能存在问题。有些问题是显性的，但更多的问题是隐性的，不易发现。如果把你阅读的研究论文与你所熟悉的研究联系起来理解，那么这些问题也不难发现。只要一直保持批评的态度，就能提高你的评价能力。有一点需要注意，批评不等于挖苦，寻找别人的弱点的同时，也要善于捕捉别人研究中的优点和独到之处，这些在你的文献综述中也许会派上用场。

功能卡片

在阅读一篇论文时，我们的兴趣一般会集中在重要内容以及它的优点和弱点上。但有一点比这些更重要，那就是这篇论文与我们自己的研究有多大关系。为此，我们必须准备第三种卡片，用来记录我们对材料的可能功用所作出的评价。这些评论需要说明这些材料将来可用在撰写论文的哪一部分。例如这篇文章是用于说明自己测量工具的设计呢？还是用于描述理论框架呢？还是用于研究结果的讨论呢？这样的评价不需要写成连贯的段落，甚至完整的句子也没有必要，只要自己能看懂、便于归档就行。根据我们的经验，有了上述 3 种卡片的参考，我们对文献的利用就会更加科学，更加有效。

小结

阅读文献是每个研究者在整个研究过程中必须经历的一个阶段。在应用语言学和相关领域里有各种不同的文献来源，如索引、学术刊物、著作以及未出版的论文等。我们可以通过人工搜索或计算机搜索的手段确立需要阅读的资料。在开始阅读文献之前，先要建立一份意向性阅读书目，然后划出阅读材料的范围。接着要做的是确定哪些材料需要详细阅读。最后，制作 3 套卡片，记录阅读过程中所综合的信息。

讨论题

1. 根据图书馆的资源，列出对你自己阅读文献有用的学术刊物。
2. 假设你的论文是研究“中国英语专业学生的冒险观念和行为”，用滚

雪球的方法，建立一份该研究的意向性阅读书目。

3. 从有影响的国际核心刊物中选出两篇论文，判别文献综述中的背景文献、中景文献和前景文献。

4. 阅读一篇自选论文，并在阅读过程中制作3种卡片。

第五章　选择研究设计

完成选题、定题的任务后，下一步就是选择合适的研究设计。一般说来，在社会科学领域里，包括应用语言学，研究设计有两大类别：定量研究和定性研究。在这一章，我们首先简要介绍定性和定量两种不同的设计，然后讨论研究问题和设计之间的联系，最后讨论与选择设计相关的一些重要问题。

第一节　定量与定性

这一节首先介绍什么是定量设计、什么是定性设计，然后指出这两种研究设计之间的主要区别，接着描述这两种设计的历史与发展，最后讨论定量与定性设计相结合的好处。

1. 定义

定量研究与定性研究有许多不同之处，因此我们很难用一两句话给这两种设计下定义。为了把问题说得简单明了一点，我们先看看 Punch 的定义（1998：4）：

> 定量研究是实证研究，其数据形式是数字。
> 定性研究是实证研究，其数据形式不是数字。

按照 Punch 的定义，定量研究和定性研究都是实证性研究（empirical research），但它们的数据形式不同，因此，区别这两种研究的最好和最简单的办法是检查所收集的数据是否以数字的形式出现。还有一些研究者，如 Hughes & Tight（1996），Tashakkori & Teddlie（1998），也建议用数字—文字对照（numeric-narrative contrast）作为区别两种研究的主要标准，但他们承认这样区别在某种程度上过于简单了。请看下面两个数字

数据的例子：

例 1

你想了解学生的词汇量，于是设计了一项测试。你从大纲中随机抽取了 100 个单词，测试要求学生给出每个英语单词的中文意思。满分为 100 分，一个单词一分。试卷的总得分表示学生的词汇量。这里学生的试卷得分是以数字形式出现的，因此是定量数据。

例 2

假设你想了解学生在提高自己的词汇量的过程中使用什么样的策略。你设计了一份调查问卷，问卷中包含 20 个有关使用策略的陈述，如“我很留心老师和同学使用的生词”，“我会利用各种机会使用刚刚学过的词汇”。每个陈述下面有 3 个选项，学生只允许选择其中一个。这 3 个选项构成一个 3 分量表：经常—有时—很少。问卷回收后，将学生的回答转换成数字：用“1”表示“很少”，用“2”表示“有时”，用“3”表示“经常”。最后将所有的回答变成数字，并把所有答案的得分相加。这里总分的高低就表示学生运用词汇策略的程度。因此，策略的使用被转换成了定量数据。

与定量的数据相比，定性的数据大多数情况下是用文字叙述。比如，调查学生扩大词汇的策略也可以与 20 个学生逐个面谈，要求他们描述自己如何扩大自己词汇量的。这是个没有确定答案的问题，学生一定会给出各种各样的回答。我们如果把学生的回答逐字逐句记录下来，我们收集的数据就是文字数据。

根据 Punch 的看法，数据形式是区分定量和定性两种研究设计的主要标准。事实上，除了数字—文字的主要差别外，它们还有其他许多重要差别，这些差别影响研究过程的各个方面。下面我们将描述这些差异。

2. 更多的差异

表 5.1 根据 Malhotra（1993：159）的表格修改而成。它列出了定量设计和定性设计之间的其他主要区别。

	定 量	定 性
目 标	检验确定的变量；验证假设	确定变量；产生假设
问 题	数据收集前是明确的	在研究过程中逐步具体化
样 本	大样本	小样本
数据收集	固定的程序结构	灵活和动态的程序结构
数据分析	统计分析	定性分析为主，也可用统计分析
结 果	有广泛的应用性	缺乏广泛的应用性

表 5.1：定量和定性设计之间的区别

Malhotra 认为，定量和定性设计的研究目标不同。前者的目标是调查科研文献中已经确定的变量，是检验有关这些变量的种种假设；后者的目的是确定需进一步研究的变量，形成未来研究可检验的假设。换言之，定量研究多用于证实以前的研究成果，而定性研究则多用于探索性研究。

用于定量设计的研究问题事先是具体明确的，是在收集数据之前提出的。明确了研究问题，实际上等于确立了选择研究对象、数据收集和数据分析的程序。因此研究者着手研究时一切都是具体明确的。相反，用于定性设计的研究问题在收集数据前比较宽泛，但随着研究的深入会逐步聚焦，所以，在定性研究中数据收集的聚焦不断变化，数据分析的方式也会时常变换。

定量研究需要大样本才能满足统计上的要求，而定性研究通常只能涉及小样本，其原因是定性研究耗时费力，大样本难以承受。就数据收集而言，定量数据的收集程序通常比定性数据的收集程序固定、刻板。相反，在定性研究中，数据收集的程序比较灵活，需要经常根据情况进行调整。

定量数据只能进行定量的统计分析，而定性数据可以选择定性或定量的分析。一般情况下，定性研究的数据以定性分析为主。既然定性数据有两种可选的分析手段，研究者须考虑在什么阶段、使用什么样的分析手段才能达到最佳目的。

由于定性研究多用于探索性的研究，涉及的案例数量不多，因此定性研究得出的结论，其应用性不如定量研究那样广泛。

3. 定量/定性研究设计的发展历程

定量和定性问题的争论由来已久，一直没有间断。回顾历史，我们可以粗略地区分为两个阶段：1）单一式设计阶段；2）混合式设计阶段（Tashakkori & Teddlie，1998）。

3.1 单一式设计阶段

从19世纪到20世纪中叶，在某个研究中，要么使用定量设计，要么使用定性设计。不过，有些研究者在定量或定性设计中尽量运用多种手段来收集数据。我们把这个阶段称之为单一设计时期。在20世纪50年代，定量设计研究处于优势，而定性研究处于边缘地位。定量研究之所以被广泛运用，主要是因为在20世纪定量研究在自然科学中取得了巨大成功。社会科学为了证明其科学地位也仿效自然科学。它们试图对其研究的对象进行量化，建立两个或更多变量之间的关系。当时，人们认为定量研究成果比语言叙述更具说服力。如果没有数字，人们就会怀疑其研究的科学性。

然而，到了60年代，定量设计开始遭到质疑，定性设计逐渐进入社会科学研究的主流。许多社会科学家认为，社会现实不同于自然现实，简单地将用于自然科学的研究设计移植到社会科学里是不妥当的。显然，我们不能用研究自然物体（如岩石、树木）的方法去研究人类和人类社会。此外，自然科学家可以孤立地去研究自然物体，但社会科学家作为社会的一部分是不能完全脱离社会的。从60年代起，定量和定性问题的争论变得越来越激烈，两种设计的支持者都坚持各自的观点，强调自身的优势，毫不相让。所以，两类研究者所从事的研究要么是定量的，要么是定性的

3.2 混合式设计阶段

自20世纪80年代，许多社会科学家开始意识到这种谁优谁劣的无谓争论是没有结果的，两大阵营之间都作出了和好的姿态。他们认识到，世界是复杂的，如果我们研究的最终目标是想全面了解这个世界，那么定量设计和定性设计都有其存在的必要。他们认为这两种设计都有各自的优势和缺陷，彼此可以互补，而不是对立。定量设计可用于大范围的研究，这类研究模式简明扼要，一目了然。定量设计的研究结果应

用性强。不足的是，定量研究通常过于简单化，生态信度*差。定性研究对现实世界能够做出比较真实的描述，能够展示现实世界的复杂性，但定性研究耗时费力，研究结果的应用性不强（Cohen & Manion，1991；Keeves & Sowden，1992）。Merton 和 Kendal 提倡两种设计的有机结合。他们说：

社会科学家已经开始放弃了界于定性和定量数据之间的虚假选择；他们更关心两者之间的结合，只有这样才能充分发挥各自的优势。（Cohen & Manion，1991：42）

Marton（1981）强调说，两种设计结合获得的结果往往能更全面、更深入地认识研究对象，与单一的设计方法相比，结果更可靠，信度更高。Markee（1994）持有相同的看法，他说："在现实中，定性和定量研究互为补充，共同创造新知识。"近年来，越来越多的研究者不再使用单一的研究设计。他们根据自己研究的需要，选择适合自己的研究设计（如 Wang，1999；Wu C.X.，1998；Hu，1998）。

4. 两种设计结合的优势

如上所述，与单一设计相比，混合设计得出的结果更具有说服力。下面是 Patton 用来说明混合设计优势的例子。

上个世纪 70 年代初，密歇根州的一些学校执行了新的教师责任制。这种新责任制的内容相当复杂，包括多种评估教师的综合系统程序，如标准化业绩测试，标准参照测试，同行之间打分，学生给教师打分，家长给教师打分，校长给教师打分，教师给自己打分等。校方对这种责任制给予了充分肯定，而教师工会和教师本人对此却持不同意见。1976 年春，Kalamazoo 教育协会在密歇根州教育协会和全美教育协会的支持下做了一项研究，调查教师对责任制的评估和态度，要求教师回答一份问卷，问卷包括多项选择题和自由问答题。从多项选择题获得的定量数据表明，几乎所有的教师"认为这个责任制是无效的，不适当的"（1990：20）。回答问卷的教师中有 70% 回答了一个自由问答题，他们的答案多达 101 页。

---

* 生态信度是指研究结果可用于现实生活的程度。为了严格控制研究中没有调查的因素，定量研究通常在实验室里进行。因此，从实验室里获得的结果很难成功地运用于真实情景。

在这种情况下，定量研究得出的结果让校方知道了教师对责任制的总体评价数字的缺点是“孤立的”，“僵硬的”，因为数字不能唤起读者的共鸣。此外，校董事会很容易将定量数据的结果看作是教师工会让校方丢脸的一种预谋。然而，校方读了几页教师本人的评价以及教师们用文字描述出责任制给他们带来的痛苦、恐惧和挫折，这时，他们了解到问题的严重性，并着手考虑解决这些问题的措施。

上述例子生动地说明了定量研究结合定性研究的优点。在这个例子中，定量研究包括一组标准化的多项选择题，这些问卷题很难设计，但教师回答起来却很容易，而且数据分析简单明了。然而，用数字表述的结果通常难以唤起人们在情感上的共鸣。相反，定性研究问题设计简单，但回答起来有些麻烦。与定量数据相比，定性数据更详细，内容更多样化，数据分析的难度也因此大得多。最为重要的是，自由问答题所获取的数据在定量设计中无法得到。

第二节　研究问题和研究设计的关系

下面我们将回答设计选择中的一个重要问题，即研究问题和研究设计之间的关系如何？

1. 是先有研究问题，还是先有研究设计？

在选择研究设计的时候，我们的出发点是什么？我们是先从研究问题入手，还是首先选定设计方案？合理的做法应该从研究问题入手。换言之，研究问题决定设计方案，反之则不行（Punch，1998；Tashakkori & Teddlie，1998）。从这个意义上讲，形成研究问题必须在选定研究方案之前完成。简单地说，应该首先想清楚我们准备研究什么，然后再考虑怎么去研究。如果连自己的研究目标都不能肯定，选择设计方案就无从谈起。

然而，有些研究者常常设计在前，问题在后。他们先定下设计方案，然后寻找适合的研究问题。这样一来，研究问题受研究设计的左右。这种本末倒置的做法会“导致极端的思维，偏狭和排斥的态度”（Johnson，1992：228）。他们会把不同的设计对立起来，或坚持一种设计

比另一种设计更好，而不考虑不同的设计在回答具体研究问题方面具有各自的优势。

实际上，研究设计本身无所谓“好”与“坏”。一种设计对于有些研究问题有效，但对另一些研究问题就不合适。每种设计都有各自的优点和缺点。关键是设计方案要适合具体的研究问题。

2. 研究问题和研究设计的匹配

不同的研究问题需要用不同的研究设计。一般说来，研究问题可分成两大类：一类需要定量研究设计来回答，另一类则需要定性研究设计来回答。一个科研项目中可能同时包含这两类研究问题。因此，研究问题和研究设计的匹配变得尤其重要。通常，研究问题中的措辞本身就暗示了应选择哪一种设计。例如，研究问题中如果出现像“变量”、“X和Y之间相关”、“影响第二语言成绩的因素”这类措辞，那么就要选择定量研究设计；如果是以“如何”、“为什么”等疑问词提出的研究问题，那么就用定性研究设计。刚刚涉足研究的学生对什么样的研究问题需要什么样的研究设计常常搞不清楚，其中一个原因就是不熟悉这些措辞的含义。

我们不妨来观察几个研究问题，看看其中的措辞能否给我们提供有关选择研究设计的信息。假设我们研究的是“学习动机”，第一个问题可能是：学生的学习动机和第二语言学习结果之间是什么关系？或学生的学习动机在多大程度上能够预测他们第二语言学习的成绩？像“关系”和“预测”这样的词语很清楚地说明，研究问题在本质上是定量的，因此，用定量的研究设计最合适，如用问卷调查的方法。第二个研究问题可能是：在第二语言课堂，使用哪一种教学方法（传统教学法还是交际法）更能促使中学生学习过去时？这个问题是对两种教学方法的有效性进行对比，因此需要用定量研究设计，如用实验的方法。第三个问题：为什么有些学生的动机比另一些学生强？显然，解释学习动机的原因很多，很难量化。这时，最好用定性的研究设计，如个案研究。

尽管某些词语能暗示我们该选择什么样的研究设计，但在许多情况下，研究问题中没有这些措辞。例如，如果研究问题是：学习动机好的学习者和学习动机差的学习者之间有什么差异？针对这样的问题，定性和定量研究设计都可以用，选择哪一种设计取决于这个领域里已经有了多少这方面的研究以及该研究的重点是什么。如果已经有了大量类似的

研究，而且该领域的研究已经很透彻，那么，最好选择定量研究来证实或反驳以前的研究成果。如果你对这两类学习者之间鲜明的细节差别感兴趣，而不是定量上的差异，那么，定性研究设计当然会更有效。

第三节 单一式研究设计还是混合式研究设计

所谓单一式研究设计，是指仅使用一种设计，这种设计要么是定量的，要么是定性的。研究者在这样的设计里可以从一种渠道收集数据，也可以从多种渠道收集数据。如果使用混合式研究设计，研究者可以采用定量和定性两种设计，但两种设计的顺序安排可以不同，在解决研究问题方面的作用也有所不同。在多数情况下，硕士论文或博士论文所作的研究一般使用混合式研究设计，因此，下面讨论的重点不在单一式研究设计上，而是在混合式研究设计上。

1. 单一式研究设计

单一式设计可以简单，也可以复杂。在简单的设计里，数据收集来源只有一个。我们先介绍简单的定量研究设计。假如某个研究者想调查调动学生学习第二语言积极性的最佳方法是什么。她要求学生回答一份问卷，问卷里列举了各种调动学生学习第二语言积极性的方法，要求学生给它们排序。这里，数据的来源只有一个，那就是学生对问卷所作出的回答。还是这项研究，你也可以使用复杂的设计。例如，除了让学生回答问卷外，研究者还可以请 10 所学校中的第二语言教师来回答问卷，让他们按照调动学生积极性的有效性将教学方法排序。如果再加上一项实验，研究可以变得更加复杂，通过实验，可以在有效程度上对调动学生学习积极性的不同方法进行比较。

同样，定性研究设计也可以由简单变为复杂。我们还是以同样的问题为例：什么是调动学生学习第二语言积极性的最佳方法？假如你采访 15 位学生，他们代表着一个班级中第二语言水平的 3 个层次。我们可以对学生进行单独采访，要求学生列出老师调动学生学习积极性的方法，然后对这些方法作出评价。如果这项研究只限于对学生的采访，那么这就是简单的定性研究设计。你也可以将设计变得复杂些，例如，你可以采访教师，或观察学生是如何对教师的方法作出反应，教师实际使用的

策略是什么。

我们来看两个例子。第一个例子是纯粹的定量研究设计，第二个是纯粹的定性研究设计。

例1：纯粹的定量研究设计

本研究旨在调查学习者因素和学习成绩之间的关系。1995年3月，有1,700名非英语专业学生参加了“学习者因素调查问卷”，他们来自黑龙江、山东和江苏3省的3所大学。1995年6月，他们参加了“大学英语四级考试”。回归分析得出的结果说明了哪些学习者因素能预测学生CET的成绩。(Wen & Wang, 1996)

例2：纯粹的定性研究设计

本研究试图调查，与第二语言水平较低的学生相比，第二语言水平较高的学生是否使用不同的阅读策略。研究对象有4名，他们是学习俄语和日语专业的在校本科生，英语是他们的第二外语。他们是本专业的优秀学生，但在英语学习方面，两个是中下水平，另外两个是中上水平。研究者选出两篇适合他们英语水平的阅读短文。每个学生在阅读短文时被要求说出心中所想。整个过程都做了录音，并转写成了文字。研究者对学生使用的策略进行了分类和介绍。最后对两组学生在使用阅读策略方面的情况进行了比较。(Lu, 1997)

2. 混合式研究设计

所谓“混合式研究设计”，顾名思义，总是包含定量和定性两种研究设计。但是，两种设计的安排方式有多种多样。从时间上，可以分为“顺序设计”和“平行设计”。所谓“顺序设计”，是指定量研究设计在定性研究设计之前使用，或相反。两种设计的顺序自然取决于研究问题，如果研究问题没有改动，这个顺序就不能改动。在平行设计中，定量设计和定性设计是独立分开的。如果人力资源允许，两个设计可以同时进行操作，也可以分开进行，因为彼此之间没有从属或依赖关系。从两种设计在回答研究问题的作用看，可以分为“平衡设计”和“不平衡设计”。在平衡设计中，定量和定性两种研究设计地位平等，而在不平衡设计中，其中一种设计占主导地位。通常，平行设计也是平衡设计，而顺序设计有可能是平衡设计，也可能是不平衡设计。将这两个参数（即时间和重要性）结合起来，那就会有多种设计（Tashakkori & Teddlie,

1998)。下面，我们将逐一介绍这些不同设计。

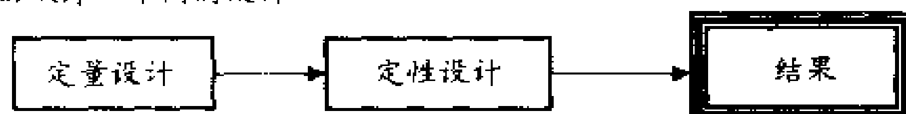
2.1 平衡的混合式设计

所谓“平衡的混合式设计”，是指定量和定性两种设计都使用，而且，这两种设计在回答研究问题时所起的作用同等重要。如图5.1所示，平衡的混合式设计有3种方案：图5.1a是平行设计，即定量设计和定性设计独立使用；图5.1b和图5.1c在顺序上不同，图5.1b的顺序是定量设计在定性设计之前，而图5.1c的顺序则相反。

a. 平行—平衡的设计



b. 顺序—平衡的设计



c. 顺序—平衡的设计

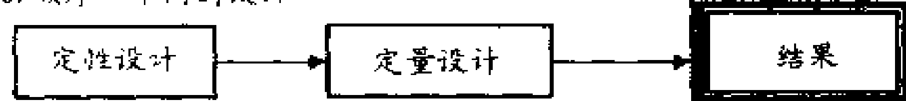


图5.1：“平衡的混合式设计”中的不同情况

下面我们通过3个实例来分别说明3种不同的“平衡的混合式设计”。

例1：平行—平衡的设计

本研究试图回答：在中国，学英语是否应该从小学一年级开始？研究对象包括学生、家长、教师和教育管理人员。该研究分为两个平行的部分：定量研究部分和定性研究部分。定量研究部分要求小学一年级学生和他们的父母分别回答两份内容不同的调查问卷。定性研究部分对小学英语教师和学校/地区的管理人员进行单独采访。所获得的两组数据同等重要，因为它们是从不同角度回答了相同的问题。

例2：顺序—平衡的设计（定量—定性）

该研究项目试图解决两个问题：(1) 第二语言写作过程在多大程度上涉及母语？(2) 在第二语言写作过程中如何利用母语？首先，100名

学生回答了一份调查问卷，问卷中设计的问题都是有关在第二语言写作过程中利用母语的情况。根据学生对问卷的回答情况，选出 15 名经常利用母语的学生参加定性研究部分，要求他们写一篇英语作文，并同时说出他们写作时的思维过程。定量研究部分得出的结果回答了第一个研究问题；定性研究部分获得的结果回答了第二个问题。

例 3：顺序—平衡的设计（定性—定量）

Gu（1997）试图调查：（1）非英语专业的本科生使用什么学习策略？(2) 这些策略的使用在多大程度上与他们的词汇学习相关？针对第一个问题，他采用了定性研究设计，对一组代表 3 个英语水平层次的二年级的本科生进行了单独采访。根据访谈的结果，设计了一份用于定量分析的调查问卷，并因此回答了第二个研究问题。

2.2 不平衡—顺序混合式设计

在不平衡—顺序混合式设计中，研究者使用定量和定性两种研究设计，但两种设计在回答研究问题时的作用不同。在图 5.2a 中，定量研究结果起主导作用，而定性的研究结果用来补充说明定量研究结果。在图 5.2b 中，定性设计比定量设计更重要，因为定性研究的结果回答了主要的研究问题。

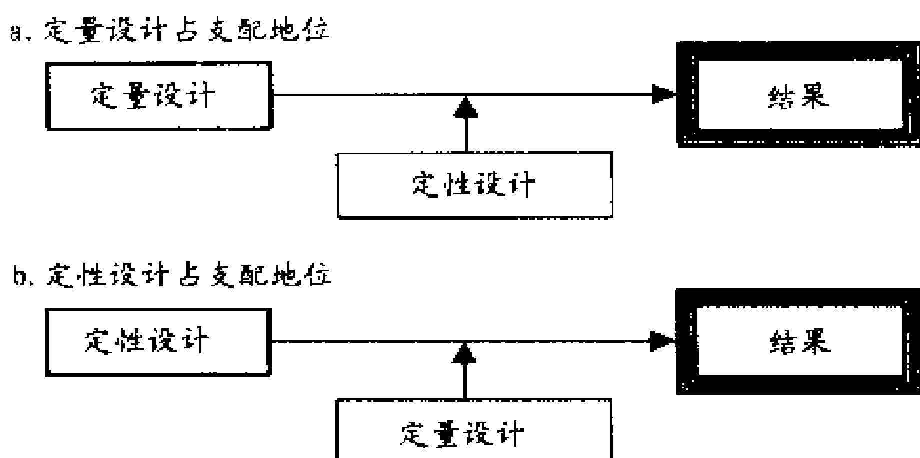


图 5.2：“不平衡的混合式设计”中的两种情况

我们来看下面两个例子：

例 1：定量设计占支配地位

Wen（1993）使用了定量和定性两种研究设计对学习者可控变量和

英语成绩之间的关系进行了研究。这项研究的定量部分要求研究对象回答一份关于学习者因素的调查问卷。调查结果向我们展示了学习者英语成绩受学习者可控因素影响程度的总体情况。这些结果帮助我们缩小了定性部分的研究范围。具体地说，定量部分的数据结果表明，就高级英语学习者而言，值得进一步研究的领域有词汇策略、母语回避策略、容忍歧义策略、管理策略以及观念和策略选择之间的关系等。然而，从定量数据中获得的结果过于宽泛和简单，因为定量研究并没有说明成功学习者和不成功学习者之间的差异。没有定性的研究部分，对统计数据的解释就会有凭空推测之嫌。定性部分的研究对象有 5 名成功学习者和 5 名不成功学习者。得出的数据揭示了有关成功学习者和不成功学习者使用策略情况的大量信息，表明学习者可控因素与英语学习成绩之间具有更加复杂、更加现实的关系。但是，如果没有定量研究提供总体研究框架，那么就很难缩小定性部分的研究范围。

例 2：定性设计占支配地位

该项目研究成功学习者与不成功学习者在冒险方面的差异。研究者首先根据过去的英语测试成绩从英语专业二年级学生中选出 16 名学生，然后对他们进行逐一采访。研究者发现了许多有趣的差异，并从这些差异中抽象出与第二语言产出有关的 4 种显著差异，以便通过调查研究来验证这 4 种差异。于是，研究者设计了一份调查问卷，调查对象有 100 名学生。

笔者经常鼓励研究生在他们的毕业论文中使用混合式研究设计，但在个别情况下，单一式设计也可以，这主要取决于研究问题。

第四节 设计分类的复杂性

不少学者（Punch，1998；Hughes & Tight，1996；Tashakkori & Teddlie，1998）认为，定量和定性研究设计之间泾渭分明，很容易区分。按照他们的分类，我们可以说，数据以数字形式出现的设计是定量的，否则就是定性的。但是，实际情形要复杂得多。下面我们用一个例子来说明设计分类的复杂性，以及解决这个问题的方法。

下面的例子选自一个研究生的论文。

Chen (2000)* 的研究试图讨论下面3个问题：

1) 中国孩子在记忆英语生词的拼写时使用策略的频率有多高?
2) 善学孩子与不善学孩子在采用词汇策略方面有无差异?
3) 学习者因素（年龄和学习英语的时间）对使用词汇策略有何影响?

这项研究的受试者是100名小学生，研究者要求他们在记忆20个生词的过程中说出他们的想法和感受。研究者反复阅读了学生的口头陈述，找出了孩子们在记忆生词拼写过程中使用的10类策略：强记，形式联想，依据熟悉词汇，依据音标，音节切分，构词法，发音规则，视觉框架，汉语音素和任意切分。现有的文献中没有这些类别的分类，这些类别是在研究者通过分析数据，归纳总结出来的，而不是她研究之前预先设想好的。通过反复检验，她确信，这10个类别的策略在有声思维数据中有充分依据。然后她数出每个类别在学生口头陈述中出现的次数，用独立样本 t 检验检查善学儿童和不善学儿童在使用词汇策略方面是否有差异。最后，她用相关分析检查策略的使用与两个学习者因素之间的关系。在结果与讨论部分，她根据从数据分析中获得的定量结果逐一回答了研究问题。具体地说，她用频次表回答了第一个研究问题；用独立样本 t 检验的结果回答了第二个问题；第三个问题的回答是一组相关系数。

上述研究属于定量设计还是定性设计？按照Punch给出的定义，我们得说，这是定性研究，因为有声思维的数据是文字形式而不是数字形式。然而，这项研究的所有研究问题都是以数字来回答的。许多研究者靠直觉就能感觉到，将这类研究只划为定性研究是不能接受的。我们再来仔细观察一下这项研究：

步骤一：收集数据

学生口头陈述的文字记录本质上不是数字形式→定性数据

步骤二：找出策略类别

研究者反复阅读口头陈述的文字记录，找出了10个策略类别→定性分析

步骤三：数出使用各个策略类别的频次

* 出于叙述的方便，Chen的论文有些地方做了改动。

研究者数出每个类别在学生口头陈述中出现的次数→定量分析

步骤四：比较善学者和不善学者使用策略的平均频次

算出每个策略使用的平均频次，并对善学者和不善学者进行比较→定量分析

步骤五：检查策略的使用与两个学习者因素之间的关系

分别算出每个策略出现频次与年龄因素和学习时间因素之间的相关性→定量分析

毫无疑问，原始数据是定性的，最初的分析也是定性分析方法，但是后来的一系列的分析都是根据统计数据的计算。更重要的是，对研究问题的回答也是用数字表示的。很显然，简单地评价 Chen 的研究是定性研究或是定量研究就不够确切。

从以上例子的分析，我们可以看出，仅仅依靠数据的形式来判断一种设计是定性的还是定量的是不合适的。我们还要考虑到数据的分析。我们建议，不要用定量或定性这样宽泛的术语描述一项研究设计，而是将数据的形式和数据分析的方法区分开来。这样，某项研究可能会有定量的数据或定性的数据，或兼而有之，数据的分析方法可以是定量的，也可以是定性的，或兼而有之。定量数据通常使用定量分析的方法处理，而定性的数据处理方法有更多的选择余地，也由于这个缘故，收集的数据如果不是数字形式，研究设计的分类就会比较复杂。

小结

研究设计有两大类别：定量研究和定性研究。有些研究者将数据的形式作为区分这两种对立设计的主要依据，即数据是数字的还是文字叙述的。除了这个特征之外，它们之间还有许多其他差异。这些差异表现在研究目标、研究问题、样本大小、数据收集、数据分析及其结果上。一般来说，研究设计在历史上经历了两个阶段：单一式设计阶段（从 19 世纪到 20 世纪中叶）和混合式设计阶段（从 20 世纪 80 年代到现在）。研究设计的选择主要取决于研究问题，研究问题中的措辞常常能提示你选择什么样的研究设计。混合式设计又有多种变体。从时序上分，可以有顺序设计和平行设计。从两种设计在回答研究问题的作用来分，又可以分成平衡设计和不平衡设计。结合“时间”和“重要性”两个维度，混合式设计又可以分为平行—平衡的设计和顺序—平衡的设计（定量—

定性或定性—定量），或不平衡—顺序设计（定量设计占支配地位或定性设计占支配地位）。本章最后分析了设计分类的复杂性，建议不要使用定量或定性这样宽泛的字眼描述研究设计，最好从数据形式与数据分析两个维度来对研究设计进行分类。

讨论题

1. 区别定量研究设计和定性研究设计的主要依据是什么？
2. 用一个例子说明将定量设计和定性设计结合的好处。
3. 研究问题和研究设计之间有何关系？
4. 定量研究设计和定性研究设计的结合有哪些形式？
5. 图 5.1b 和 5.2a 中的设计有什么区别？
6. 图 5.1c 和 5.2b 中的设计有什么区别？
7. 从国际学术刊物中找出相应的例子说明平衡混合式设计中的变化形式和不平衡混合式设计中的变化形式。（注意：可分组进行，各小组中每人负责一个例子。）

第六章　问卷调查

本章介绍如何进行问卷调查。问卷调查主要用于收集定量数据*。本章首先对问卷调查进行简要描述，接着对数据收集的程序进行详细阐述。

第一节　问卷调查简介

问卷调查通常是描述性的，牵涉的样本比较大。有的描述用于变量特点的调查，有的描述则用于变量之间关系的调查。如果研究旨在描写两个变量之间的关系，这样的研究就是个相关设计。然而，越来越多的人已意识到就两个变量来研究某一现象显得过于简单。事实上，诸多现象都是多个变量共同作用的结果。因此，更为常见的设计是多变量设计，即研究多个自变量与多个因变量之间的关系。问卷调查可以是横向的或共时的，即在同一时间从不同的研究对象那儿收集数据；问卷调查也可以是纵向的或历时的，即在不同时间从相同的研究对象那儿收集数据。

完成一项问卷调查，必须完成3项任务：1. 设计数据收集的研究工具；2. 选择研究对象；3. 收集数据。

第二节　问卷设计

问卷是用来广泛收集定量数据的工具之一。和观察、访谈相比，问卷调查花时少，费用低，但设计一份高质量的问卷却耗时费力。问卷设

* 本章不讨论收集定性数据的问卷。

计的最大困难是缺乏严谨的理论指导，因为我们要研究的问题千差万别，所以很难回答和描述什么是最佳的调查问卷。这里我们只能给出一些指导原则。虽然这些原则会帮助问卷设计者避免犯一些重大错误，但真正好的问卷来自于周密的思索和丰富的实践。下面我们对高质量问卷的标准，设计问卷的过程以及问卷的分级测量技巧逐一进行介绍。

1. 高质量问卷的标准

高质量问卷至少要符合两个标准：(1) 内部效度高；(2) 外观样式正规。

1.1 内部效度高

所谓内部效度高是指问卷题必须测量想要研究的变量。这一标准看似有些多余，因为没有人会傻到不去测量他想研究的变量。实际上，由于考虑不周、技巧不熟练、经验不足，这些都可能导致疏漏或不当。研究新手常犯的通病就是内部效度低。我们不妨来看几个例子。有个本科生想研究高中生在阅读过程中是否使用母语。她设计了以下 3 道问卷题：

1）为了帮助我了解某个英语故事发生的背景，我常看中文的参考书。

2）我在业余时间看英语小说、报纸、杂志等。

3）我只用英汉双解词典。

假使学生在回答第一条时，选择了“这个句子完全或几乎完全适合我的情况”。这个回答意味着他们在二语阅读时使用母语吗？我们很难得出明确的结论。至于第二条，经常读英语并不能保证在二语阅读时避免使用母语。第三条可能在研究中根本不起作用，原因是差不多所有的高中生都使用英汉双解词典。再说，即使有少数学生使用英英词典，答案还是不能直接给我们有关在二语阅读时使用母语的信息。

有位硕士研究生想研究学生在说二语时的焦虑程度。她的问卷中有这样 3 条：

1）上课时，我尽量避免回答问题。

2）我一般坐在教室的拐角处，这样老师叫不到我回答问题。

3）我不喜欢在大庭广众之下讲话。

如果有个学生承认以上这 3 条都符合他的情况，你能说这个学生在

讲二语时焦虑感很高吗？显然不能。这是因为上课避免回答问题可能有多种原因：除了焦虑外，还可能是他上课前没有好好准备，他本身二语能力差，他性格内向，等等。由此可见，上面3条除了测量焦虑外，还测量了其他因素，因此这些问卷题的效度就很成问题。

为了检验自己设计的问卷，你不妨反复问自己："学生对某一题的回答意味着什么？答案只有一个解释，还是多个解释？"如果只有一个解释直接与研究的变量有关，那么这一条的效度就很高，否则就要修改或删除。

1.2 问卷外观样式规整

问卷的外观样式会影响调查对象回答问卷的态度，从而影响他们回答的质量。例如：如果问卷使用的纸质太差，或问卷的排版比较难看，调查对象会认为这问卷不重要，结果就不认真回答。一份高质量的问卷会注意以下几个问题：

第一，如果问卷超过5页，不要用回形针或订书针把它们订在一起，最好把它们装成小册子，并采用双面印刷形式，这样看上去比较正规。第二，同一道题不要跨两页，因为题目分开在两页上，调查对象会误认为题目已经结束。第三，不要为了节省空间，把许多题目挤在一起。这样的问卷容易导致调查对象错误回答问题，也容易给调查对象造成问卷很复杂的印象，降低他们回答问卷的积极性，最终导致回答率的降低。第四，问卷的纸张质量要好，打印要清楚。外观样式要正规、专业。这样不仅可以减少调查对象的疲劳和厌烦，而且可以提高回答率。

2. 问卷设计过程

问卷设计包括以下几个步骤：1）选择实施问卷调查的方式；2）明确收集数据的类别；3）决定数据的处理方式，是单个处理还是分类处理；4）确定每题的内容及措辞；5）选择问卷题目的结构；6）将问卷题排序；7）决定问卷的格式；8）进行小范围的试测（Malhotra，1993）。

2.1 选择实施问卷调查的方式

实施问卷调查的方式有两种：直接发放问卷和邮寄问卷。第一种是调查组织者直接把问卷发放给调查对象，另一种是调查组织者把问卷邮寄给调查对象。调查组织者可以是研究人员本人，也可以是协助调查的其他人。用第一种方式调查，一组人可以一起回答同一份问卷。如果问

题不清楚，调查组织者可当面解答；调查组织者还可以检查调查对象是否回答了所有问题。这样答卷质量会有所提高，数据缺损率也会明显降低。

如果调查对象分布在各个城市，甚至各个国家，我们就只能使用第二种方式实施问卷调查。这样，研究人员无法控制总的答卷率和数据的缺损率。即使答卷如期寄回，研究人员既无法确信答卷对象为何人，也无法确定答卷完成的具体环境。

由于实施问卷调查的方式不一样，答卷方式的说明和问卷题的数量就不一样。一般说来，邮寄问卷的长度要相对短一些，问卷题数要相对少，答卷方式的说明要更加简洁明了。很显然，我们不能奢望调查对象用一两个小时来回答邮寄的问卷，除非他们能获得相应的报酬。

这两种问卷调查方式各有利弊。究竟选择哪一种方式取决于其费用、方便程度和研究问题的性质。我们认为还是当面发放问卷的方式比较好，因为答卷者如果有疑问，可以当面及时澄清，此外，调查者还可以检查每份答卷的完整性。

2.2 明确收集数据的类型

确定了实施问卷调查的方式以后，我们就应该确定哪些类型的数据是我们想要收集的。一般说来，有 5 类信息：

(a) 经历/行为（调查对象所为）；

(b) 看法（调查对象所想）；

(c) 情绪感受（调查对象所感）；

(d) 知识/能力（调查对象所知和调查对象所能）；

(e) 个人简况（调查对象的出生日期、年龄、性别及婚姻状况）。例如：Wen（1993）在研究二语学习者可控因素与英语学习成绩之间的关系时，使用的问卷包括三类信息：第一部分：个人简况；第二部分：看法，即观念；第三部分：行为，即学习策略。下面是该问卷的部分内容（pp. 294 - 302）。

语言学习者因素问卷

下面的问卷旨在研究中国学生学英语的方式。所有这些问题都经过精心选择，并涉及到英语学习的各个重要方面。请根据你的看法或学习经历如实回答每一题。每题的答案无对错之分。所收集的数据只用作研

究，我们负责保密。

第一部分 个人简况

1. 姓名________出生年月__________性别________

2. 高考英语入学成绩______________________

3. 高考语文入学成绩______________________

第二部分 观念

下面是人们对外语学习的一些看法。这些看法无对错之分。我们只对你的看法感兴趣。请大家根据每个数字所代表的含义选出其中一个数字填在句末的括号里，所填的数字一定要能如实代表你自己的看法。

1=我坚决不同意这个看法

2=我不同意这个看法

3=对这个看法我没有明确的答案

4=我同意这个看法

5=我非常同意这个看法

201 很好地计划自己的学习时间是学好外语的重要保证。 ( )

202 学外语需要花工夫。 ( )

203 学外语重复练习很重要。 ( )

第三部分 学习策略

下面是人们常用的一些外语学习策略，这些策略无对错之分。请根据数字所代表的意思，选择其中一个填在每个句末的括号里，所选数字一定要能如实地描述你的学习情况。请记住，填写时要根据自己的实际做法，而不是你的想法或其他人的做法。

1=这个句子完全或几乎完全不适合我的情况。

2=这个句子通常不适合我的情况。

3=这个句子有时适合我的情况。

4=这个句子通常适合我的情况。

5=这个句子完全或几乎完全适合我的情况。

301 当阅读课文时，我争取弄懂课文的每一处。 ( )

302 我背诵课文。 ( )

303 我课外自己对自己说英语。 ( )

我们在实际操作过程中，可以把以上5种信息与时间框架（过去、

现在和将来）联系起来。例如：关于某个行为的题目可指调查对象过去上中学时的行为，也可指现在上大学时的行为，还可指未来读研究生的行为。

初涉研究的人往往会把观念题和行为题混为一谈。其实，这里还是有规律可循的。一般说来，观念题的结构是“我认为/想……”。例如：“我认为学好外语需要花功夫”这句话就是想搞清调查对象的看法或观点，而不是其行为。而“我花了好多功夫学习外语”这句话是想调查学习者的行为，而不是其看法与观点。同样，如果调查对象把观念看成行为或把行为看成观念，数据的效度就值得怀疑。

2.3 决定问卷数据的处理方式

一旦决定了实施问卷的方式和收集哪一类信息，接下来就需要决定如何处理问卷数据：是单个题目处理，还是分类处理？如果数据分析的基本单位是单个题目，比如说逐一描写每个题目的频率和平均分，这样的问卷称作为单个题目问卷。编写这种问卷相对来说比较简单、容易，因为它不涉及建立抽象类别。然而，这类问卷的价值极其有限，因为其结果缺乏概括性和广泛的应用性。

设计以抽象类别为基础的问卷，一方面需要建立类别，另一方面需要确定每个类别包括哪几个问卷题。确立抽象类别的方法有两种：“自上而下”和“自下而上”。所谓自上而下的方法是指在设计具体的题目之前，根据逻辑和现行理论先确定抽象类别。所谓自下而上的方法是指想到什么题目，就把这题目写下来，然后再进行分类。很显然，设计以类别为基础的问卷对于设计者来说要求很高，比较富有挑战性。然而，所有的硕士研究生都要掌握这个技能。例如：为了研究英语专业二年级学生外语学习的观念和策略，Wen设计了一份以类别为基础的问卷，具体类别见表6.1。

表中传统学习方法与非传统的学习方法以Stern（1975，1983，1992）提出的二语学习中的3大争端为理论基础，即1）母语的作用；2）外语学习是有意识的行为还是无意识的行为；3）学习者交际时，是应注意语言形式的准确性还是应注意语言的流利性。每个争端都是一个连续体。以第一大争端为例，即：二语学习中是否使用母语，有人极力主张在外语教学中使用母语作媒介语，鼓励学习者在学习外语时使用翻译，有人则坚决反对借助母语学习外语，鼓励学习者在学习外语时用外语思

维。围绕这一争端，Wen分别设计了依赖母语观念和回避母语观念；依赖母语策略和回避母语策略。

方法	传统的方法	非传统的方法
观念	形式操练观念 依赖母语观念 准确性观念	功能操练观念 回避母语观念 流利性观念
策略	形式操练策略 依赖母语策略 准确性策略	功能操练策略 回避母语策略 流利性策略

表6.1：观念与策略的分类

这里"方法"被定义为观念与策略的结合，其理论基础是Marton和Saljo（1976）提出的教育心理学。他们主张学生的学习行为受其学习观念的影响，换句话说，学生的学习观念会影响他们学习策略的选择。

Chen（1996）为了弄清教师的态度在多大程度上影响学生的焦虑，使用了以类别为基础的问卷调查高中生的焦虑。在最初设计的时候，他并没有确立焦虑类别，只是想到什么就把什么写下来，然后再进行分类。经过反复推敲，他决定把它们分为两类：环境焦虑和内在焦虑。环境焦虑是指由二语学习环境引起的焦虑，例如课堂气氛和教师对学生的态度；内在焦虑是指学生个性引起的焦虑。

2.4 决定问卷题的内容及措辞

一旦决定了问卷信息的处理方式，我们就得开始考虑每个问卷题的内容。问卷题的来源很多，不应该闭门独自在那里苦思冥想。首先可以看看前人在这一方面所做的研究。如果你想研究大学生二语学习的策略，可以参照Oxford设计的《语言学习策略问卷调查表》（1990）或Wen的《学习者因素问卷调查表》（1993）。也可以先对几个大学生进行访谈，了解他们二语学习的情况。访谈所收集的资料可成为你撰写问卷题的重要依据。

如果问卷题涉及好几个抽象类别，你必须考虑题目是否覆盖所有的类别，每一类里面的问卷题是不是数量适当。为了回答这两个问题，必须设计一个表格，一列是类别，另一列是问卷的题目号（参见表6.2）。

类别名称	题目号
形式操练观念	2，4，5，7，9
意思操练观念	1，3，6，8，10
形式操练策略	11，13，16，18，20
意思操练策略	12，14，15，17，19

表 6.2：检查每类问卷题目表

反复阅读每类的问卷题目，确保这些题目与该类别的名称相吻合。如果发现有些题目不适合某一类，应该进行修改或设计新的题目。在这一阶段遇到这样的问题不足为怪，但不能置之不理。最后，必须仔细检查题目的措辞，如果题目的措辞不好，调查对象可能会不理解或误解。在这样的情况下，答案中可能会有好多缺值或误导。为了避免措辞不当，下面提 4 点建议（Malhotra，1993；Bernard，1994）。

避免导向性问题

避免导向性问题是指不要给调查对象有引导性暗示的问题。例如：

不受人欢迎的精读课应该改革。

坚决不同意	不同意	不明确	同意	坚决同意
1	2	3	4	5

很显然，上面这个问题给调查对象以“坚决同意”的暗示，这是因为“不受人欢迎”是一个带有偏见的词语。

使用普通词语

在设计题目的时候应使用普通词语，词语的难易程度要适合调查对象的语言水平。假如调查对象是小学生，问卷题所用的词语必须在他们所学的范围内，否则，他们不能回答，或所给的回答效度很低。

使用正反两方面的题目

在问卷中，正反两方面的题目要保持平衡，换句话说，正反两方面的题目数要相当。例如：文秋芳、吴学顺（1998）在研究学生对使用语言实验室教口语的态度时，设计了这样的问卷，其中有 3 道题是从正面设计的，5 道题是从反面设计的（后面带星号★的，表示是从反面设计）。

1=这句话完全或几乎完全不适合我的情况。

2=这句话通常不适合我的情况。

3=这句话有时适合我的情况。

4=这句话通常适合我的情况。

5=这句话完全或几乎完全适合我的情况。

1. 我喜欢在语言实验室上口语课。
2. 我不清楚为什么要在语言实验室上口语课。★
3. 在语言实验室，这么多人同时对着话筒讲话，噪音太大，使我无法集中思想思考要说的内容。★
4. 我不喜欢老师监听我和别人的交谈。★
5. 我认为在语言实验室上口语课，便于把听和说有机地结合在一起。
6. 我认为在口语课上听力练习应该尽可能的少。★
7. 我认为在语言实验室内通过话筒与别人交谈缺乏交际的真实性。★
8. 我认为在语言实验室内对着话筒讲话就好比平时打电话一样。

从上面例子，我们发现反面的句子含有明显的否定标记，如：第二题中的“不清楚”，第三题中的“无法”，第四题中的“不喜欢”，有时有些句子使用不明显标记，如：第六题中的“尽可能的少”和第七题中的“缺乏”。

我们为什么要这样做？因为研究表明调查对象的回答受到正反两方面句子的影响。

翻译要谨慎

翻译要谨慎的建议是针对跨文化研究而提出的。在跨文化研究中，一般要牵涉到两种或两种以上不同的语言。有时，不同语言之间表面上相同的两个句子却含有不同的意思。我们怎么才能确保翻译不歪曲原义？最好的方法是回译。例如：Wen（1997）曾研究中国学生与瑞典学生的学习观念。最初的问卷用英语编写而成。他们先请人把这份问卷分别译成汉语和瑞典语，然后又请两个人把问卷回译成英语，通过反复比较最初的英文和回译的英文，他们发现了一些问题，然后再进一步地修改，直到对译文感到满意为止。

2.5 选择问题的结构

一般说来，问题分成两种形式：一种是开放式的问题或非结构式的

问题，另一种是封闭式的问题或结构式的问题。

开放式的问题指的是调查对象可以用自己的话回答问题，这些问题的答案是开放式的，难以预测。请看几个例子。

你为什么不喜欢看英文简写本小说？

你什么时候查词典？

你课后看什么？

你为什么学习英语？

从上面的例子，我们发现开放式的问题多涉及以 wh 开头的词，如：什么（what）、什么时候（when）、为什么（why）。开放式的问题一般说来适用于定性研究，不能量化，不能做定量分析。

在由封闭式问题或结构式问题组成的问卷中，研究人员给调查对象提供一整套答案供选择。换句话说，调查对象必须从所提供的选项中选择答案。封闭式的问题可能是多选题、二选题或分级题。

多选题

在多选题中，研究人员提供一整套的备选项，要求调查对象从所给的选项中选出一个或多个答案。我们来看一看 Wu, C.X.（1998）问卷中的两个例子。

1. 在日常生活中，你是如何学习文化差异的？（可选多个答案）

(a) 看英语杂志和有关西方风俗习惯的书

(b) 看英语电影和录像

(c) 主动与外国人交流

2. 你在和外国人谈话时，最注重哪一方面？（请选一个答案）

(a) 语言形式的准确性

(b) 意思清楚

(c) 讲话的流利性

(d) 讲话有礼貌

二选题

二选题一般只有两个选项，如："是"还是"不是"；"重要"还是"不重要"；"同意"还是"不同意"，等等。有时，两个选项中还加一个中性选项"不知道"。例如：

1. 课外你对自己讲英语吗？

——讲

——不讲

2. 你常看英语报纸吗?

——常看

——不常看

3. 你认为有语言天赋的人不用功可以学好外语吗?

——可以

——不可以

——不知道

分级题

分级题的备选项形成一个连续体，在这个连续体中，有一整套可供选择的选项。例如：

1. 你读英语小说消遣时猜生词的意思吗?

从不猜	偶尔猜	有时猜	经常猜	总是猜
1	2	3	4	5

2. 你认为理解英语课文的最好方法是翻译吗?

坚决不同意	不同意	不知道	同意	完全同意
1	2	3	4	5

很显然，分级题与多选题类似，两者都提供一整套的选项，然而，它们的性质不同，分级题可形成一个连续体，而多选题则不能（有关分级题的详细情况参见下一部分的“分级技巧”）。

2.6 确定题目的顺序

题目的顺序常常会影响调查对象的回答。关于编排题目的顺序，我们有下面两点建议：

1）调查对象个人简况放在前面

调查对象的个人简况通常放在调查相关变量问卷题（包括观点、行为和知识等）的前面。然而，Malhotra（1993）和 Liu（1999）则认为个人简况应该放在后面。

2）总的问题放在具体问题的前面

总的问题通常放在具体问题的前面，这样会有助于防止具体问题影响对总的问题的回答。例如：如果你想弄清楚学生在他们的阅读过程中使用什么策略，你最好先问“你怎么进行阅读的?”，然后再问“你在阅

读过程中是如何处理生词的？”。

2.7　确定问卷的格式

在问卷付印之前，需要决定问卷题目的格式和位置。这对问卷设计来说也很重要。研究表明，调查对象对前面的问题比后面的问题要更重视，把问题分成几个部分往往可提高答题率。此外，所有的问题必须编号，这样便于调查对象答题。

2.8　对问卷进行小规模的试测

一个有经验的研究人员在正式印制问卷之前，一般都要进行小规模的试测，这样可发现潜在的问题，及时予以修正。潜在的问题是多方面的，有题目的内容、措辞、顺序、难度、答卷说明和格式，等等。选择进行小规模试测的调查对象必须与正式研究的调查对象相仿。最好小规模试测的方法是通过访谈，在访谈中，研究人员可以观察调查对象的反应和态度，检查他们对问题的解释是不是与研究人员的用意相符。最后，应该对小规模试测的结果进行分析，看看题目是不是合适。如果有分类的话，看看分类是不是恰当。

第三节　问卷答案的分级

结构式问卷中的答案一般有两项或多项供选择。在数据分析中，多项问卷答案要能够进行比较、计算。为此，设计问卷时，我们必须对问卷答案进行分级。所谓分级就是将问卷答案放在同一连续体上。例如：你可能把人们对学习二语的态度标为消极的、中性的和积极的。或者你把使用策略频率分为经常、有时和偶尔。这里的“消极的—中性的—积极的”和“经常—有时—偶尔”是两个连续体。应用语言学中使用的分级方法包括比较分级和非比较分级两种。

1. 比较分级

比较分级就是将供选择的答案进行比较。比较的项目有时是两项，例如：你是喜欢在公共场合讲话还是喜欢在小组里讲话。有时是多项，例如：你如何提高阅读技能？（请根据频率高低将下面选项排序）

1）看英语杂志

2）看英语报纸

3）看简写本小说

4）看原版小说

两项比较称为配对比较分级，多项比较称为序列分级（见图 6.1）。比较分级只能产生等级数据，为我们提供相对的解释。

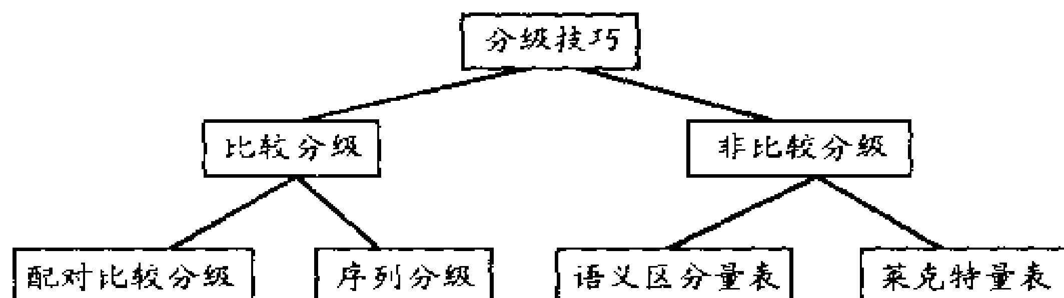


图 6.1：分级方法的分类（根据 Malhotra，1993：282 改编）

1.1 配对比较分级

采用配对比较分级，问卷题只给调查对象两个选择。例如：

1）讲英语时，你比较注重下面哪一方面？

（1）准确性

（2）流利性

2）你什么时候会感到更紧张？

（1）跟外国人讲英语时

（2）跟英语老师讲英语时

3）你什么时候想多讲英语？

（1）在小组里

（2）在两人组成的对子里

1.2 序列分级

与配对比较分级不同，序列分级时，每个问卷题必须提供 3 个或 3 个以上的选择，同时要标明排序的标准。例如：

1）你为什么要学英语？（根据重要性将下列选项排序）

（1）对学外语感兴趣

（2）想毕业后找个高薪的工作

（3）打算到国外学习

（4）想了解更多的外国文化

2）你什么时候想更加努力学英语？

（1）英语考试得了高分的时候

（2）在课堂上受到英语老师表扬的时候

（3）在英语演讲比赛中获奖的时候

根据 Malhotra（1993）的看法，采用比较分级有3点长处：首先，调查对象必须对两个或多个选项的细微差别进行比较；其次，由于比较的标准相同，调查对象答卷中存在的差异一目了然；再次，这样的分级不需要任何理论假设。

2. 非比较分级

比较分级需要将两个或多个选项进行比较、排序，其结果属于等级测量。与比较分级不同的是，采用非比较分级时，问卷题有3个或3个以上的选项，这些选项在一个连续体上是等距的。调查对象只要挑选其中的一个，其结果属于等距测量。下面介绍的语义区分量表和莱克特量表就是最常用的非比较分级（参见图6.1）。

2.1 语义区分量表（Semantic differential scale）

语义区分量表采用两个语义相反的词语，中间有5到7个刻度，调查对象从中选择一个以表明自己的态度或行为。例如：Gardner & Lambert（1972）采用语义区分量表研究父母对孩子学习法语的鼓励程度。

你学法语时，父母给你多少鼓励？

一点儿没有____：____：____：____：____：____：____很多

他们还用类似的方法研究调查学生对法国人的印象。

有趣____：____：____：____：____：____：____乏味

有偏见____：____：____：____：____：____：____无偏见

勇敢____：____：____：____：____：____：____胆怯

漂亮____：____：____：____：____：____：____难看

丰富多彩____：____：____：____：____：____：____单调无味

友好____：____：____：____：____：____：____敌意

诚实____：____：____：____：____：____：____狡猾

愚蠢____：____：____：____：____：____：____聪明

这儿有一点必须强调的是，否定的形容词或短语不应该固定在量表的一边。调换一对反义词语的左右位置可以在一定程度上控制某些调查

对象的答卷倾向。根据我们的经验，有些答卷者有肯定或否定倾向，即倾向于选择正面答案或负面答案。一旦否定和肯定词语的位置不固定，调查对象就必须逐一阅读答案，做出选择。否则他们可以不看题目，把答案直接标在一边即可。

2.2 莱克特量表（Likert scale）

莱克特首次在问卷中采用了等级测量，莱克特量表由此而得名。该量表一般采用3、4、5或7个选项，每个选项都有文字描述，每相邻两个选项之间的距离相等。例如让调查对象在一个4分量表上评估他们学英语的原因：1=不重要；2=有点重要；3=重要；4=非常重要。为了进行分析，每个回答都给予了相应的分值。使用这一方法决定每个调查对象的总分时，一定要使用前后一致的评分程序，这样高分或低分的含义才能一致。例如对否定句来说，同意表示否定，而对肯定句来说，同意表示肯定。因此对肯定句"完全同意"的回答，与对否定句"坚决不同意"的回答应该得到同样的分值。

第四节 选择调查对象

一般说来，我们不可能去研究总体。其原因显而易见："你不可能研究每个地方的每一个人所做的每一件事"（Miles & Huberman，1994：27）。常用的方法是从总体中选择一个样本进行研究，并希望能把样本中的发现应用到总体中。

选择调查对象与问卷设计同样重要，同样有难度。如果调查对象没有选择好，收集数据的质量肯定有问题。如果所选调查对象不肯合作，甚至都不露面，那么数据收集就难以进行。这里，我们先讨论选择调查对象的标准程序，然后重点讨论一些具体操作事宜。

1. 随机抽样方法

随机抽样方法有3种：简单随机抽样、系统随机抽样和分层随机抽样。

1.1 简单随机抽样

简单随机抽样是最基本的抽样方法，是系统随机抽样和分层随机抽

样的基础。其操作方法简单、方便，因此应用范围最广。简单随机抽样时，第一步需要把对象总体的每个个体赋予编号，从 1 到 N（N 表示总体的数目），接着使用随机数字表在其中选出 n 个作为样本（随机数字表见附录 5）。所谓随机数字表，就是表中的每个数字可能被抽取的概率完全相同。在抽样过程中，比对象总数大的数无效，已经被选用过的数亦无效。

下面是一个简化的随机数字表。

91567	42595	27958	30134	04024	86385	29880	99730
46503	18584	18845	49618	02304	51038	20655	58727
34914	63976	88720	82765	34476	17032	87589	40836
57491	16703	23167	49323	45021	33132	12544	41035
30405	83946	23792	14422	15059	45799	22716	19792
09983	74353	68668	30429	70735	25499	16631	35006
85900	07119	97336	71048	08178	77233	13916	47564

表 6.3：随机数字简表（Black，1992：256）

假如你需要运用简单随机抽样从 20 所大学中挑选出 10 所。首先，必须给每所大学编号（见表 6.4）。使用几位数作为第一个标号数取决于总体数量的位数，换句话说，如果总数的位数是两位数，就以 01 开始；如果是三位数，就以 001 开始。

01 北京大学	11 辽宁师范大学
02 清华大学	12 四川大学
03 北京师范大学	13 武汉大学
04 复旦大学	14 南开大学
05 华东师范大学	15 中山大学
06 南京师范大学	16 河南大学
07 上海大学	17 汕头大学
08 东南大学	18 浙江大学
09 南京大学	19 山东大学
10 河海大学	20 厦门大学

表 6.4：20 所大学的编号

将 20 所学校编好号后，我们就要从随机数字表中进行抽样。随机数字表中的数字 5 个一组，每次看几位呢？这里总体数量是 20 所学校，因此每次查看的数字应该是两位。表 6.3 的第一行的第一个两位数是 91，这个数字大于 20，超出范围，不能进入样本。第二个两位数是 56，接着下来是 74、25、95、27、95、83。所有这些数都超出范围，不能进入样本。83 后面的两位数是 01，这个数可以用。跳过 34 后，是 04 和 02，这两个数可以进入样本。接下来第一行的数字是 48、63、85、29 和 88，均大于样本数 20。下面一个可用的数字是 09。剩下的 97 和 30 都不能用。用同样的方法，我们到了第二行。第二行所有的数字是 46、50、31、85、84、18、84、54、96、18、02、30、45、10、38、20、65、55、87、27。除了 18、10 和 20 这 3 个以外，其他数字均不能使用，它们要么是大于 20，要么是前面已使用过了，如 02。至此，我们已有了 01、02、04、09、10、18、20。我们继续使用同样的方法，又得到了 3 个数字：08、11、03。下面的这些大学是通过简单随机抽样选中的。

01 北京大学	09 南京大学
02 清华大学	10 河海大学
03 北京师范大学	11 辽宁大学
04 复旦大学	18 浙江大学
08 东南大学	20 厦门大学

表 6.5：通过简单随机抽样所选中的大学

在上面例子中，我们采用了逐行检查随机数字表中的数字。如果采用逐列检查的方法也完全可以，这纯粹是个人偏爱。但一旦定下来，就必须前后一致，不能中途更换方法。此外，除了超出范围和重复出现的数字，不允许跳过任何数字。

1.2 系统随机抽样

如果总体很大，而所要的样本很少，就不宜使用简单随机抽样，因为所选的样本并不是均匀分布在总体当中。假如一个总体是 900 名学生，想从中随机抽样 30 名学生。按照简单随机抽样，只要到 900 的一半，就可以选出 30 名。为了避免这个问题，我们使用系统随机抽样的方法，以确保样本均匀分布在总体之中。和简单随机抽样一样，系统随机抽样也需要给所有的个体编号。此外，还需要计算抽样间距。下面的公式可

用来计算抽样间距，

$$抽样间距=\frac{总体的总数}{样本的大小}$$

我们如何从900个人中选30个人呢？首先要计算出抽样间距，即拿900除以30等于30。接下来每隔30选出一个。从哪个数字开始呢？这里需用随机数字表选出一个1到30之间的数作为起点。根据表6.3，第一个数字是915比900大，因此不能进入样本。下一个数字是674可以进入样本，接下来的一个数字是704（即674+30），这样依次下去，得到的抽样号码为734、764、794、824、854、884。884之后有点麻烦，原因是，884再加30就等于914，而914超过了900。怎么办呢？解决方法很简单：拿914减去900，所得到的14就是第九个数字，这样再依次类推44、74等等。最后被选中的30个数字是：

674、704、734、764、794、824、854、884、14、44、74、104、134、164、194、224、254、284、314、344、374、404、434、464、494、524、554、584、614、644

如果计算出的抽样间距带有小数，就将小数4舍5入变成整数。例如抽样间距是29.4，就变成29。

1.3　分层随机抽样

假如一所部属重点大学有900名理科专业一年级的学生，其中有90名女生。如果想从中选出30名学生作为研究对象，且男女生人数相等。这时，如果仍使用系统随机抽样，女生数可能只占到10%。在这样的情况下，可以用分层随机抽样的方法，以确保样本的性别比例相同。首先，把学生分成男生组和女生组，然后再用系统随机抽样从每个组里获取所需的样本。具体地说，从90名女生中随机选出15名，从810名男生中也随机选出15名。

如果想要30名学生的样本能代表原来总体的性别分布，即10%是女生，90%是男生，同样采用分层随机抽样。首先，把学生分成男生组和女生组。然后，再用系统随机抽样，从90名女生中随机选出3名，从810名男生中随机选出27名。在应用语言学中分层随机抽样应用非常广泛，例如：研究人员可能想获得代表二语水平的不同层次的样本，或代表不同家庭背景的样本，或代表不同母语背景的样本。在分层随机抽样中，首先要把总体分组，然后再进行随机抽样。由此可见，分层随机抽

样要比前面两种随机抽样复杂得多。

问卷调查的对象不能低于 30 人。这是统计分析的最低要求。下面我们还要讨论：选择什么样本？要从样本中收集几次数据？

1.4 方便抽样

在实际研究中我们时常不可能进行严格的随机抽样。退而求其次，只好采用方便抽样，即“能找到谁就用谁”，用这样的方法得到的样本有其先天的不足之处，用于探索性研究和小规模的问卷试测问题不大，但必须承认，这样获取的样本用于正式研究可能存在问题。

2. 选择调查对象时的注意事项

调查对象的选择必须放在整个研究计划中通盘考虑，其中特别要关注的是研究问题。具体地说，选择调查对象需要考虑以下 3 个问题：

1）此项研究选择多少调查对象？为什么？

2）如何选择调查对象？这些调查对象在多大程度上能代表总体？为什么？

3）在此项研究中，调查对象要参与几次？为什么？

2.1 是大样本还是小样本？

如果问卷调查是研究的主体部分，并且研究的主要目的是显示二语教学总的模式和总的发展趋势，那么，就需要选择大样本。然而，如果问卷调查只是研究的一部分，并且研究是探索性的，那么，就选择相对的小样本。人们不禁要问多少样本为大样本，多少样本为小样本。其实，样本的大小是相对而言的。根据我们对统计要求的理解，问卷调查的最小样本不少于 30，大样本不低于 500。如果研究需要不同组别之间的比较，每组至少需要 30 人。

2.2 是随机样本还是方便样本？

当研究必须要具有代表性时，就得使用标准化的随机抽样方法来获得随机样本：或简单随机抽样，或系统随机抽样，或分层随机抽样。作为研究生，由于受到许多条件的限制，往往不能随机抽样，只能使用方便样本。对于方便样本，还是需要决定选择样本的程序。例如：你的样本是在自愿的基础上获得？还是根据他们的二语水平获得？还是通过老师的推荐获得？很显然，如果自愿参加，数据收集时，与研究人员合作就不会出现问题，不足之处只是缺乏代表性，因为自愿的学生往往学习

动机比较强。相比较而言，完整的班级作为方便样本比自愿报名得到的样本代表性要强。此外，还必须考虑研究对象的统计特点，如：多少男生？多少女生？如果所选的研究对象是同一性别，要解释其原因。再如：有些研究对学习者的年龄差异感兴趣，这样，年龄又成了选择研究对象的重要参数。

2.3 是纵向研究还是横向研究？

如果一项研究旨在发现某些变量的变化或揭示其发展模式，那么，需要在不同的时间点上考察研究对象。这样研究对象就形成了纵向样本。例如：你想研究学生经过高等教育后，学习英语的动机有没有发生变化，你可以用同样的问卷调查同一批学生两次，第一次是他们刚入校的时候，第二次是在他们快要毕业的时候。最后把两次问卷的结果进行比较，从而看出 4 年的变化。有时研究人员无法连续几年跟踪同一批研究对象，这时可以分别选择一年级和四年级的横向样本进行研究。还以刚才的学习动机为例，你可以同时选择一年级和四年级学生回答问卷，然后再比较这两组人的问卷结果。这两组学生之间的差异可以看成一到四年级的变化。横向研究的好处是比较经济，省时省力，其弊端是，造成两组学生之间的差异，接受 4 年的高等教育只是原因之一，另外的原因可能还包括两组学生之间存在的其他差异。

第五节 实施问卷调查

很多情况下，我们愿意花大量的时间和精力去设计问卷，却舍不得花同样的时间和精力去收集数据。实际上，设计问卷与实施问卷对于问卷数据的质量同样重要。下面是我们对实施问卷的 3 点建议。

1. 以专业人员的身份接近研究对象

以专业人员的身份接近研究对象，这一点很重要。当研究人员把调查对象集中到教室并准备回答问卷之前，需要对这项研究的有关情况做一简单介绍，以争取他们的合作。

介绍由两部分组成。第一部分简要讲解该项目的背景，其中包括项目的重要性、研究目的、数据收集方式；第二部分讲答卷的具体要求和

需要注意的事项。在讲解研究目的时，要讲究技巧。假如研究的目的是考察母语的使用与二语水平之间的关系，如果你直截了当地把这一目的如实告诉调查对象，很可能会影响他们的答题选择。如果我们换一种说法：本次研究需要了解母语在二语学习中的作用，前面的问题就不会产生。除了介绍研究背景以外，研究人员还要向调查对象当面承诺，他们的名字不会在研究报告中出现，他们的答卷除了供研究人员使用以外，不作任何其他用处。

关于答卷的要求一定要讲得具体。尽管问卷上的要求写得清清楚楚，最好还是要把这些要求再重复一遍，因为有些粗心性急的人常常不看书面的要求就动笔答题。要告诉他们答题时如有困难和问题，请及时举手提问。临交卷时，还要提醒他们要回答所有的题目，避免不必要的疏漏。最后不要忘记向调查对象表示感谢。如果财力允许的话，赠送他们一支笔或一个讲义夹以表谢意。

2. 尽可能由研究者直接发放问卷

问卷调查实施的方式对数据的质量有很大的影响。笔者建议：第一、如果有可能，直接发放问卷，而不要邮寄问卷；第二、如果有可能，研究人员亲自实施。

如果调查规模比较大，调查对象分散在不同城市的几所大学里，研究人员就难以亲临每所大学。比较常见的做法是，事先培训协助调查的人员，以确保问卷调查的要求一致。

3. 选择良好的环境和适当的时间

在什么地方、什么时间实施问卷调查也影响数据的质量。例如：在教室里回答问卷的质量与在宿舍里回答问卷的质量不同；靠近吃午饭的时候回答问卷的质量与刚吃过早饭后回答问卷的质量不同。很显然，为了使收集的数据有信度和效度，一个安静的环境和适当的时间非常重要。根据我们的经验，问卷调查的最佳时间是上课的时候。这样做必须征得上课教师的同意。有他们的老师在场，又在正常上课的时候，学生肯定会合作，认真对待问卷调查。

小结

问卷调查一般牵涉到的样本较大，使用封闭式的问卷来收集定量分析的数据。一份好的问卷必须要有很高的内部效度和正规的外观样式。设计问卷时，必须要关注以下几点：1）确定实施问卷调查的方式；2）确定收集数据的种类；3）确定处理数据的方式；4）确定每条问卷题目的内容；5）选择问卷题目的结构和措辞；6）按适当的顺序编排问卷题；7）确定问卷的格式；8）在小范围试测问卷。在问卷调查中，使用分级方式来建立一个连续体。应用语言学中使用的分级方式包括比较分级和非比较分级两种。比较分级对选择进行直接比较。当比较分级在两个项目中进行时，称为配对比较分级，否则就叫序列分级。在非比较分级中，调查对象不需要对多项选择进行比较。由此而产生的数据一般为等距分级。有两种非比较分级量表：语义区分量表和莱克特量表。语义区分量表使用两个语义相反的词，中间有 5 到 7 个刻度，让调查对象选一个刻度表示其态度或行为。莱克特量表含有形成一个连续体的多项选项，每个选项用文字表述，每两个选项之间的距离相等。如果调查对象的人数少于总体的人数，必须进行随机抽样。随机抽样的方式有 3 种：简单随机抽样、系统随机抽样和分层随机抽样。实施问卷调查时，应该以专业人员的身份接近研究对象，同时，要在安静舒适的环境中进行，选择的时间要恰当。

讨论题

1. 一份好问卷的基本标准是什么？
2. 从国际学术期刊上选一份问卷，仔细检查问卷题的内部效度。
3. 简要描述问卷的设计程序。
4. 运用不同的分级技巧，设计问卷题目。
5. 在选择调查对象时，应考虑什么？
6. 在下面的研究中，最好使用哪种随机抽样？
   1）某研究人员想把 106 名高二学生分为两组进行比较研究。
   2）1999 年，南京大学外国语学院录取了 180 名新生，其中英语专业 70 名，法语专业 25 名，德语专业 24 名，日语专业 26 名，俄语专业 20 名，西班牙语专业 15 名。我们如何从中选择具有代表性的样本 30 名？

3）某所大学有1,000名理科学生和860名文科学生。在理科学生中，80%是男生，20%是女生 在文科学生中，52%是男生，48%是女生。我们怎样才能得到100名文理科学生均等，男女生均等的随机样本？

第七章　实验研究

和问卷调查一样，实验研究也属于定量研究。与问卷调查相比，实验研究更难付诸实施，但它对建立因果关系更为有效。本章首先要解释什么是实验研究，接着讨论实验研究中的两个重要概念：因果关系和效度。最后描述各类实验研究，以及进行实验研究的程序。

第一节　什么是实验研究？

实验最初用于自然科学领域。我们都有在物理实验室和化学实验室做实验的经历。那些实验给人留下印象最深的是试管、化学药品、电线和其他一些材料。那么，究竟什么是社会科学领域的实验呢？在社会科学的实验研究中，研究人员在控制某些干扰变量以后，对一个或多个自变量进行调控处理，然后测量这些自变量对因变量的影响。实验研究包括下面这些基本要素：

1．对一个或多个自变量进行调控处理；
2．比较至少两组人或两种情况；
3．测量受调控的自变量对一个或多个因变量的影响；
4．采取措施控制干扰变量对因变量的影响。

两组实验对象，一组叫实验组，另一组叫对照组。如果要对这两组进行比较，他们之间除了实验处理不同以外，其他各方面都应该完全相同。换句话说，实验组接受某种实验处理，而对照组没有。一旦实验完毕，研究者必须对因变量进行测量。如果两组之间存在差异，差异就归结于是否接受某种实验处理。

问卷调查与实验研究的最本质的区别是前者的环境没有人为的加工，而后者的环境接受了人工调控。典型的调控是让研究对象在规定的条件下接受某种实验处理。研究人员所进行的调控越多，所推出的因果

关系就越有说服力。

我们来看实验研究的几个例子。

例 1：两组人之间的比较

Bejarano（1987）想研究合作学习法是否比传统的全班学习法更为有效。这项研究在以色列进行，参加实验的是 665 名 7 年级学生。经过随机分配将他们分为实验班和对照班，实验班改用合作学习法，对照班依然使用全班学习法。经过 4 个半月的教学，所有学生参加了听力理解和阅读理解测试。研究结果发现就听力理解训练而言，合作学习法优于全班学习法，但对阅读理解训练来说，全班学习法优于合作学习法。

在上述例子中，研究对象接受了不同的学习法，受到了不同的实验处理。实验结果是学生在听力理解测试和阅读理解测试中的成绩。实验班使用的是合作学习法，而对照班依然使用全班学习法。为了控制干扰变量的影响，实验组和对照组的人员分配采用了随机抽样。

例 2：两种情况之间的比较

Pica、Young 和 Doughty 3 个人在 1987 年进行了一项研究，旨在发现通过商谈交流（negotiation）是否能促进二语理解。他们进行了一个小规模的实验，对两种语言输入的方式进行比较。16 名中级水平的成人参加了这项研究。参加者要求在两种不同的情形下，通过在木板上放东西来检查他们的语言理解能力。第一种情形是：由英语本族语者宣读指令，这些指令的语言经过加工后，变得简单、易懂，例如简化了句法结构、解释了指令中困难词语、重复了指令中的实义词。但是，学生在听指令时不允许与宣读指令的人进行语言交流。第二种情形是：指令仍旧由同一位英语本族语者宣读．与第一种情况不同的是，指令的语言原封不动，没有经过任何简化，但指令宣读者鼓励实验对象与他进行交流，以便理解指令的内容。研究结果表明，学生放东西的准确率在第二种情形下比在第一种情形下高。由此，研究者推断商谈交流可促进二语理解。

在上述实例中，实验处理牵涉到同一组研究对象接受指令的两种不同情形。测量实验结果的方法是要求受试者根据对指令的理解在木板上放东西。研究结果为受试者在两种不同条件下理解指令的正确率。由于只涉及到一组研究对象，因此许多干扰变量已经得到了控制。

第二节 因果关系

因果关系这一概念对于实验研究非常重要。下面，我们首先讨论因果关系的涵义，接着讨论因果关系的条件。

1. 因果关系的涵义

科学家对“甲引起乙”这句话有多种解释：

甲可能是引起乙变化的多种原因之一；

甲可以解释乙的变化；

甲与乙之间存在因果关系，尽管不能证明，但可以推断。

上述解释与我们日常对因果关系的理解有些不同（Malhotra，1993）。普通人往往从绝对意义上看待因果关系，把原因看成为结果的决定因素。因此一说到“甲引起乙”，人们就会认为甲是导致乙发生的惟一原因，并且相信这一事实可以观察，也可以验证。

很明显，科学家对因果关系的解释更加适合应用语言学研究。二语教学非常复杂，有多种变量交织在一起。这其中的因果关系往往属于概率倾向、不可能被直接观察，也不可能被直接验证。因此，即便存在着真正的因果关系，也可能没有被人们发现。

2. 因果关系的条件

一般说来，因果关系必须同时满足3个条件：因果之间的时间先后顺序、因果之间的必要联系、因果关系之间没有其他可能存在的原因（Maxim，1999；Punch，1998）。这3个条件，缺少其中的任何一个都不能构成因果关系。

时间先后对因果关系的确立非常重要。如果说甲是乙的因，那么甲必须发生在乙的前面，而不是乙在甲的前面。这一观点似乎很清楚，而且能解释不少因果关系。例如：语能可解释二语成绩的差异，母语水平影响二语成绩。从时间的角度看，语能和母语水平都发生在二语学习前面。然而，如果仅仅依赖这一条件来确定因果关系，很容易出现错误。例如：夏季总是在春季的后面，冬季总是在秋季的后面，但我们不能因此就说：春是夏的因，或秋是冬的因。因此，时间先后这一个条件不足

以确定因果关系。

因果关系的第二个条件是必要联系，如果两个变量构成因果关系，这两个变量之间必须存在联系。换句话说，甲的变化应该与乙的变化有联系，或者说这两个变量必须是共变。问题是在许多共变的变量之中，很难分清哪个是因、哪个是果。例如：二语学习动机与二语成绩就是这样一对变量。如果我们能够分清这两个共变变量之间的时间顺序，因果关系的确立就容易得多。这道理很简单，发生在后面的事情，不可能成为发生在前面的事情的原因。即使满足了“共变”与“时间顺序”这两个条件，我们仍不能百分之百地确定这两个变量之间的因果关系，因为我们所研究的变量可能不是引起变化的真正原因。

第三个条件是能够排除可能存在的其他原因，也就是说，研究人员应该能够解释为什么那些可能存在的原因不可能发生。这里有两种方法可以运用。一种方法是人工控制（physical control），即通过随机抽样来确定实验组和对照组，这样可以消减两组之间的差异。另一种方法是统计控制（statistical control），即在分析数据的过程中，通过统计的方法来控制干扰变量的影响。

第三节　实验的效度

评价实验研究的基本标准是效度。什么是实验研究的效度？哪些因素可能会影响效度？怎么保证实验研究的效度？下面我们来逐一讨论这几个问题。

1. 内部效度和外部效度

在实验研究中，有内部效度和外部效度两种。所谓内部效度是指所得到的因果关系在多大程度上能够得到圆满的解释。设想在实验研究中，研究人员成功地创造了一个环境。在这个环境中，只有自变量对因变量发生作用，其他变量都被有效控制。这时因变量的变化完全归结于实验的作用。这样的实验研究，我们就认为效度很高。换句话说，有内部效度的设计可以确保是自变量而不是其他变量引起因变量的变化。那些不属于研究范围的变量称作为外部变量或干扰变量。控制干扰变量是取得内部效度的关键。

外部效度是指研究结果的推广性。Bracht & Glass（1968）指出推广性的两个主要方面：研究对象和环境。“研究对象的效度”是指某一研究结果在多大程度上可应用到其他研究对象身上。“环境的效度”是指某一研究结果在多大程度上可推广到其他环境中去。

2. 影响内部效度的因素

影响内部效度的因素很多，有些与环境有关；有些与研究对象有关；有些与变量的测量方法有关；有些与研究对象的选择有关；有些与实验条件有关。下面我们来逐一描述。

2.1 与环境有关的因素

环境因素是指影响实验进行的外部情景因素。例如：Brown（1988）曾在中国从事过研究。作为该项研究的一部分，他比较两组听力理解测试的成绩。第一组进行测试时，教室窗户开着，外面非常安静。然而，几个小时以后，当第二组测试时，突然外面有拖拉机发动机的隆隆声，即使把窗户关起来，声音还是很大。很显然，这两组之间的听力成绩比较变得没有任何意义。原因很简单，因为第二组学生测试的环境不如第一组。除了噪音以外，还有很多其他环境因素需要控制，例如温度的高低、时间的早晚、光线的强弱、通风的程度和座位的舒适程度等都可能影响实验的效度。否则，这些环境因素和自变量一起发生作用，必然会搅乱自变量和因变量的关系。

2.2 与研究对象有关的因素

与研究对象有关的因素包括受试者的成熟、受试者的流失和霍桑效应。

受试者的成熟是指研究对象随着时光的流逝而自然产生的变化。这种变化与自变量的作用没有直接关系。在二语学习中，随着二语学习者年龄的增长，生活经验的积累，知识面的扩大，他们变得逐渐成熟起来。这种生理、心理的成熟必然促进二语水平的提高。因此，如果一项研究持续好几个月，甚至好几年，研究结果必然要受到成熟因素的影响。

受试者的流失是指在实验过程中受试者中途退出。退出的原因多种多样。例如：实验令人乏味、耗时太多、要求太高，或实验的时间对研究对象不方便。作为研究者，一般很难确定中途退出的受试者是否与留

下来的人具有相同特点 设想流失者属于学习动力不足、成绩差的一类，实验结果是否一定由自变量的影响所引起就要受到质疑。

根据 Brown（1988）的报告，Mayo、Roethlisberger 和 Dickson 曾在美国西部电气公司芝加哥霍桑分公司进行实验。尽管时间不同，工作条件不同，但只要他们一到场，生产效率就提高。经分析，他们认为生产效率的提高很可能是因为工人想取悦研究人员。后来人们把在社会科学里，受试者为取悦研究人员而产生的积极表现称为“霍桑效应”。例如：在研究任务教学法是否有效时，研究者告诉受试者他们即将参加的实验是为了检验新方法是否比传统方法好。在这种情况下很可能受试者认为自己能被选中参加实验班是一种荣誉，因此好好表现，以不辜负教师的希望。由于这样的原因，研究结束时，实验班的学生可能比对照班的学生成绩好，这时就很难说清楚这一结果是由于实施了任务教学法，还是霍桑效应的结果。

2.3 与测量有关的因素

另一些可能会影响研究内部效度的因素是测量方法。例如：我们使用测试、问卷、访谈和其他手段来测量一个或多个变量。测量变量的方式可能会影响内部效度，这其中的重要原因有两个：测试效应和测量的不稳定性。

所谓测试效应是指由测试本身所引起的效应。产生该效应的情况多种多样，下面谈到的是其中的三种情况：

第一，测试效应可能会发生在收集数据的过程中。例如：如果前测与后测之间的间隔时间太短，研究对象就有可能还记得前测所给的答案。为了表明他们的观点一致，有些研究对象在后测时尽可能选与前测相同的答案。也有些研究对象在后测时揣摩出研究人员的用意，尽可能选与前测不同的答案。这样，后测答题的效度就会受到影响。

第二，前测的题目可能使研究对象意识到要研究的变量。结果，他们对变量敏感度的提高导致了学习效果的改进。例如：Brown（1988）曾想研究如何有效地让学生学会使用缩略形式。在教学之前，实验组与对照组参加了缩略形式的前测。接着，在教学过程中，实验组采用明示教学法教授缩略形式，而对照组没有。研究结束时，两组再次参加缩略形式测试，结果实验组明显比对照组考得好。然而，Brown 无法确定实验组的高分是仅仅因为采用了明示教学法所致，还是研究对象前测时对缩

略形式的高度敏感也起了很重要的作用。

第三，当同一份语言水平测试反复使用，以确定学生取得的进步时，也会产生测试效应。后测的进步可能不仅仅是由教学所致。例如：如果同一份语法测试考了两次，在第一次考试中犯的一些错误在第二次考试时可能避免，原因是学生自己能够从错误中学习新东西。这样，研究结果就比较模糊，难以解释。

除了测试效应以外，研究工具的不一致性或不稳定性也会影响内部效度。设想对同一组学生实施两次不同的测试，一次是前测，一次是后测，目的是看看他们的成绩有没有进步。如果后测比前测难得多，那么很难说清楚两次测试的差异是实验的失败，还是由于两次测试的难度没有掌握好。

2.4　与选择研究对象有关的因素

研究对象的选择可能也会影响研究结果。例如研究者选择两组学生来比较任务教学法和传统教学法的教学效果。一组学生在研究开始前，就已经使用了某种教学法，他们对这种教学法就比较熟悉，也比较适应。而另一组要接受的是一种全新的教学法。这样，一开始这两组就不在同一起跑线上。因此，两组学生后测的差异可能不是教学方法的不同所致，而是由于事先存在的差异。有时，选择的研究对象是自愿者。一般情况下，自愿者都是学习态度比较认真、学习动机比较强的人。从这些自愿者身上得到的研究结果往往很难解释，因为我们不清楚研究对象的动机强度对研究结果产生了多大影响。

2.5　与实验实施有关的因素

影响实验研究内部效度的因素可能与实验的实施有关。例如实验时间的长短就是一个重要因素。设想你想比较学生相互批改与教师批改对二语写作能力发展的影响。如果你的实验只有 4 个星期，结果发现实验组的写作成绩并不优于对照组。根据这样的结果，你便得出以下结论：学生相互批改并不比教师批改更为有效。我们的问题是：二语写作能力的发展是一个缓慢渐变的过程，4 个星期怎么能够显出明显进步呢？因此，这项研究的内部效度令人怀疑。

除了实验的时间长短之外，实验实施的方式也可能影响实验的内部效度。例如：你想研究“加注法”是否对提高二语写作能力有效。所谓加注法，就是要求学生在自己的作文中通过加注的方法说明自己写作中

的困难，这样教师看到学生的加注后，就能根据提出的困难给以针对性的帮助。除了“加注法”以外，实验组与对照组的教学内容都相同。整个实验持续了一年。结果实验组在后测中比对照组作文成绩高得多。看起来，我们可以把实验组的好成绩归于加注法的实施。但这里我们忽视了加注需要时间这一重要因素。为了加注，实验组肯定每次作文所花的时间比对照组要多得多。这样一来，我们就很难说清楚实验组的作文水平提高是因为加注法，还是因为花了更多的时间，还是两者兼而有之。

最后，实验研究的内部效度还可能受到参与该实验的实施者个体的影响。例如：你请两个教师进行实验，一个教师教实验班，另一个教对照班。这两位教师的情况怎么样呢？他们是不是有相同的教学经历？他们的二语水平是不是相同？假如教实验班的教师经验比较丰富，而教对照班的教师相对经验不足，最终的结果实验班比对照班好，我们就很难说实验班的好成绩是因为实验所致，还是因为教师的丰富教学经验所致。

3. 影响外部效度的因素

做实验时，研究者总希望自己的研究结果能应用到其他学生或其他环境中去。然而，与实验对象和环境有关的因素可能会影响其应用的范围。

研究对象的选择可能影响外部效度。例如：如果你想了解中国中学生的英语阅读策略，就不能选择外国语学校的学生作为样本。如果你希望研究结果应用范围广的话，在选择样本时就要谨慎。设想你正在研究如何有效教授/r/和/l/这两个音素。如果你这个实验放在设备先进的语言实验室里进行，整个程序严格控制，你就不能把实验室内的研究结果等同于普通教室内的研究结果，其原因是语言实验室的环境与现实生活中的普通教室环境很不一样。

4. 两种效度之间的关系

如果研究人员希望自己的研究结果既对自己有用，又对同一研究领域的其他研究人员有用，那么，内部效度和外部效度都很重要。然而，在这两种效度之间有个权衡。为了提高内部效度，研究人员尽可能按照实验室的条件来设计控制研究程序。这样做，内部效度是提高了，但外部效度就有所降低，因为实验室的条件与真实环境相差很远。同样，提

高外部效度的同时很可能会降低内部效度。在自然环境中进行实验能够增加外部效度，但在自然环境中许多干扰变量与研究变量共存，内部效度就没有保证。因此，研究者应该尽可能在两种效度之间取得平衡，以求得两种效度的最大化。然而，在现实中要在两种效度中保持平衡并非易事。如果难以取得平衡，内部效度应该优先考虑，因为，没有内部效度，外部效度便无从谈起。

5. 控制干扰变量

干扰变量会导致研究结果有其他的解释，它们严重威胁着实验的内部效度和外部效度。由于这个原因，干扰变量又称为混淆变量。一般有两种方法来控制干扰变量：一种是人为控制，另一种是统计控制。人为控制是指在实验设计的过程中研究者对变量加以有意控制，而统计控制是在分析数据的过程中通过某种统计方法对变量加以控制。表 7.1 是控制干扰变量的各种策略（Punch，1998：83）。

人为控制	统计控制
随机抽样	受试者分层
限制变量	部分相关
匹配分组	协方差分析

表 7.1：控制干扰变量的策略

5.1　人为控制

如果在设计实验过程中想控制干扰变量，可以使用以下这 3 个策略：随机抽样、限制变量个数和匹配分组。

随机抽样是指把实验对象随机分成实验组和对照组。尽管这样并不能保证实验组和对照组完全相同，但可以保证在这两组中干扰变量不可能对因变量产生系统影响。也就是说，即便有干扰变量存在，他们的影响在一定程度上可以相互抵消。因此，随机抽样是保证实验组与对照组等同的一个程序。然而，如果样本太小，随机抽样的效果可能不大。当然，我们可以通过测量可能的干扰变量，比较组与组之间的差异，来检查随机抽样是否有效。

限制变量就是采取措施使得某些变量恒定不变。也就是说，被限制的变量在实验中没有发生变化。例如：如果研究人员认为性别是干扰变

量，可能会影响因变量，从而影响实验的内部效度，他就可以选择同性别的研究对象来参加实验。这样，性别就不可能对因变量产生任何影响。然而，同性别的研究成果就不一定能推广到另一个性别上去。况且，限制一个变量并不能保证研究结果不受其他干扰变量的影响。因此运用限制变量的方法并不能百分之百地解决实验效度问题。

匹配分组是指在实验前，根据某些关键的背景变量，对受试者进行分组。例如：在研究阅读策略训练的效果时，以动机、二语阅读水平和性别把两组学生进行配对，然后再把他们分成实验组和对照组。匹配分组有两个明显的不足：第一，由于研究对象是根据几个特征进行配对，因此，他们在其他一些变量上可能是不相等的；第二，如果所匹配的特征与因变量不相关，那所做的匹配就是一种无效劳动。

5.2 统计控制

我们有时让自然组参加实验，既不取消两组之间的差异，也不限制变量数目。这样，两组在实验前可能存在着系统差异。如果我们已经意识到这样的差异，在分析数据时，可以采取一些措施来测量这些差异。这样，根据测量的差异，我们再来调整这些差异所造成的影响。具体地说，从自变量的影响中去除干扰变量对因变量的影响，这个过程就称为统计控制。实施统计控制时，我们可以采用下面 3 个策略：受试者分层、部分相关和协方差分析来进行统计控制。

1）所谓受试者分层（Rosenberg，1968）就是把一个干扰变量选为调节变量，然后根据调节变量将受试者重新分组，再进行数据分析。例如：如果你认为二语水平是个干扰变量，你可以把它选为调节变量，将受试者重新分组。然后数据分析分别在同水平的受试者中进行。由于所有的研究对象都在同一语言水平，经过分析后产生的自变量与因变量的关系就不可能再受二语水平的影响。

2）部分相关是把干扰变量对因变量的影响从自变量对因变量的影响中排除出去。例如：某项研究中有 3 个变量：动机、学习策略和二语学习成绩。很显然，在现实生活中，动机与学习策略一起影响二语学习成绩。如果想弄清楚学习策略在多大程度上影响二语学习成绩，我们就必须去除动机的影响。在统计分析时，我们可以运用部分相关分析来达到这一目的。

3）协方差分析是指在进行主要分析之前，先把协变量的影响从因变

量中排除掉。假如你的实验旨在发现阅读策略培训的有效性。组成实验组和对照组的学生是两个自然班。对照组正常上阅读课，而实验组除了正常的上课，还要教他们阅读策略。两个组都参加了阅读理解的前测和后测。由于这两个组不是随机抽样，他们在实验前就有明显差异。为了控制实验前阅读能力这一干扰变量，你对阅读理解的分数进行了调整，以消除在前测分数中的差异。经过调整后，再对两组进行正式比较。

以上 3 个策略在本质上有一个共性，那就是想方设法消除干扰变量对因变量的影响，以提高实验的效度。

第四节 实验研究的种类

实验研究可分为 3 类：前实验研究、准实验研究和真实验研究。就研究人员的控制程度而言，它们是一个连续体。前实验研究所受控制最小，真实验研究所受控制最大，准实验研究介于两者之间。区分 3 种实验研究的关键是实验组与对照组在多大程度上相似。

	前实验研究	准实验研究	真实验研究
不同的设计	单组后测设计 X　O_1	非等值组后测设计 实验组：X　O_1 对照组：　O_2	等值组后测设计 实验组：R　X　O_1 对照组：R　　O_2
	单组前后测设计 O_1　X　O_2	非等值组前后测设计 实验组：O_1　X　O_2 对照组：O_3　　O_4	等值组前后测设计 实验组：R O_1 X O_2 对照组：R O_3　O_4
不同的特征	没有对照组 没有随机抽样	有对照组 没有随机抽样	有对照组 有随机抽样

表 7.2：不同种类的实验研究

使实验组与对照组相同的最好办法是对受试者进行随机抽样。真实验研究采用随机抽样的方法分配实验组与对照组，而前实验研究和准实验研究都不具备这个特征。准实验研究尽管没有随机抽样，但实验组与对照组之间的差异还可以识别、测量出来。前实验研究所受控制最少，它既没有对照组，也没有随机抽样。

1. 前实验研究

前实验研究的特点是既没有对照组，也没有随机抽样。它的具体设计有两种：单组后测设计和单组前后测设计。

1.1 单组后测设计

单组后测设计可以用下列符号表示：

X　　　O_1

（X 代表实验处理，O_1 代表实验后的因变量测量。）

举例：某项实验要比较新方法与传统方法的优劣。某个研究者假设一种新教学方法比传统的教学方法更能提高托福的成绩。她连续 5 年举办了托福培训班，都是使用传统的教学方法。5 年中学生的托福成绩都一一作了记载。每年的平均成绩大约在 540 分到 575 分之间波动，5 年的总平均为 567 分。今年她采用了新的教学方法。培训班一结束，所有的学员都参加了托福考试，他们的平均分是 585 分。她把今年的平均分与以往 5 年的平均分做了比较，发现今年的学员比以前的学员考得好。

在上述实验中，研究者把今年样本的平均值与一个已知值或总体参数进行了比较。通过这一比较，我们可以看出这个组与已知值的差异。有时我们可能会怀疑某一次学生的考试成绩是否优异。这时我们就可以把学生的考试成绩与以往学生的成绩做比较。

这种实验设计的应用范围极其有限，因为只有当我们能够得到过去对因变量的测试成绩，并且用以测量因变量的测试具有权威性时，这种实验才有价值。如果缺少这两个条件，这种实验设计的缺陷是显而易见的。第一，不可能把实验条件下因变量的情况与没有实验情况下因变量的情况进行比较；第二，因变量的情况可能受到多种干扰变量的影响，如受试者的成熟程度、受试者的流失以及受试者的选择。缺少对这些干扰变量的控制将会削弱实验研究的内部效度。鉴于这些原因，单组后测研究比较适合探索性的研究，一般不用于正式研究。

1.2 单组前后测设计

单组前后测设计可以用下列符号表示：

O_1　　　X　　　O_2

（X 代表实验处理，O_1 代表前测，O_2 代表后测）

举例：某项研究要考察阅读策略培训的有效性。

为了弄清楚阅读策略培训是否能提高学生的阅读理解能力，某研究人员决定在她班上进行实验。在培训开始前，她对学生的阅读理解能力进行了测试。10个星期以后，她用与前测同样难度的试题，对这些学生进行了后测。然后，把两次测试的平均分进行了比较，结果发现学生的后测成绩比前测成绩好。

在这个设计中，没有对照组进行比较，但同一组学生参加了两次测试。按照这样的设计，这个组在进行实验（X）之前，参加了前测（O_1）。实验完成以后，又参加了后测（O_2）。实验效果是 $O_2 - O_1$。这一设计在自然科学里行之有效，因为在自然科学里，无生命物体能够在一段时间里保持不变。而在社会科学里，由于各种干扰变量的存在，如：受试者的成熟、测试效应、研究工具、研究对象的选择以及研究对象的流失，这一设计的效度仍旧令人怀疑。后测与前测之间的时间越长，干扰变量对因变量影响的可能性就越大，该设计的效度问题就越大。

2. 真实验研究

真实验研究与其他两种实验设计相比，最重要的区别特征是采用随机抽样的方法分配实验组与对照组。

2.1 等值组后测设计

等值组后测设计与非等值组后测设计有一个共同的特点：它们没有前测。但这两者的区别在于等值组后测设计使用了随机抽样，而非等值组后测设计没有使用随机抽样。等值组后测设计可以用下列符号表示：

实验组：	R	X	O_1
对照组：	R		O_2

（R代表随机抽样、X代表实验处理，O_1代表对实验组的后测，O_2代表对对照组的后测）

举例：某项研究试图比较多次修改法与多篇作文练习法的有效性。

为了提高二语写作水平，本次实验检验多次修改法是否比多篇作文练习法更有效。两个随机抽样的班级参加了这次实验。一个班（实验组）要对同一篇作文进行多次修改，另一个班（对照组）要练习多篇作文。具体地说，实验组写6篇作文，每篇作文修改3次；对照组写18篇作文，不进行修改。除了这一差别以外，两个班的作文课其他方面都相同。一学期结束时，对两个班的学生进行作文测试，出乎研究人员的预

料，两个班在写作能力方面没有明显差异。在这个设计中，实验效果为 O_1-O_2，也就是说，实验组测试的结果减去对照组测试的结果。尽管在实验之前，进行了随机抽样以确保这两组所有的相关特点都相同。当然这种相同是研究者自己认定的。由于没有前测，我们就不能进一步验证。因此，这种实验设计仍可能受到研究对象的选择、研究对象的流失和霍桑效应等因素的影响。尽管这种实验设计有这些局限性，但它具有一些明显的优势，例如采用了随机抽样并有对照组，只需要测试一次。

2.2 等值组前后测设计

在等值组前后测设计中，将实验对象随机抽样分成实验组和对照组，每组都进行前后测，这一设计可用下列符号表示：

实验组：	R	O_1	X	O_2
对照组：	R	O_3		O_4

（R 代表随机抽样，X 代表实验处理，O_1 和 O_3 代表实验组和对照组的前测，O_2 和 O_4 代表实验组和对照组的后测）

在这个实验设计中，如果统计学上的 O_1 和 O_3 完全相等，那么随机抽样的假设在逻辑上就能站住脚。然而这样的假设在现实中得不到百分之百的实现。如果使用前测，可进一步控制两组之间在实验前的差异。我们仍旧以多次修改作文的实验为例，假如我们对这一实验稍做改动，增加一次前测。下面是两组的前后测分数：

前测：$O_1=75$；　$O_3=80$

后测：$O_2=85$；　$O_4=85$

根据上面前后测的成绩，我们很容易就可以计算出实验效果：

$(O_2-O_1)-(O_4-O_3)=(85-75)-(85-80)=5$。但如果没有前测，该实验可能就会被误认为没有效果：$(O_2-O_4=85-85=0)$。

3．准实验研究

准实验研究与前实验研究不同的是：准实验研究有对照组；准实验研究与真实验研究不同的是：准实验研究没有随机抽样。

3.1 非等值组后测设计

这一设计牵涉到两个组：实验组和对照组。实验结束时，两组均参加测试。受试者的挑选带有随意性，没有经过随机抽样程序。这一设计可以用以下符号表示：

实验组：　X　O_1
对照组：　　O_2
（X代表实验处理、O_1代表对实验组的后测，O_2代表对对照组的后测）

实验的效果为O_1-O_2。然而，由于没有随机抽样，这一实验处理的效果也可能受到了一系列干扰变量的影响，也许两组在实验前就存在差异，也许受试者的选择就带有偏见。

3.2 非等值组前后测设计

与非等值组后测相比，非等值组前后测要求有前测。这一设计可以用以下符号表示：

实验组：　O_1　X　O_2
对照组：　O_3　　O_4
（X代表实验处理，O_1和O_3分别代表实验组和对照组的前测，O_2和O_4分别代表实验组和对照组的后测）

计算实验效果的方法是：$(O_2-O_1)-(O_4-O_3)$。用这样的方法已经在某种程度上考虑了实验前的差异。然而，一个前测不可能揭示实验前所有的差异。设想某个实验对受试者性别或者个性比较敏感。这就是说，这个实验可能对某种性别或某种个性有利。然而，前测并不能反映这些差异。因此，如果没有随机抽样，干扰变量不可能完全受到控制。

毫无疑问，准实验研究不是检验假设的最好选择。只有在真实验研究无法进行时，再选择使用准实验研究。

第五节　实验研究的程序

完成一项实验研究应遵循哪些程序？按照逻辑顺序，研究人员应首先制定实验计划，然后再进入实施阶段。在制定实验计划时，研究人员就受试者的选择、前测、对受试者实施自变量的影响和后测作出一系列的决定。整个计划进行了充分论证后，研究人员开始实施计划。请记住在实施过程中还要进行监察、调整和记录。下面就制定计划和实施计划两个阶段逐一讨论。

1. 制定实验计划

在制定实验计划阶段，研究人员需要回答以下几个问题：

1）需要多少实验对象？如何选择？

2）如果有前测，如何进行前测？

3）如何对受试者施加自变量的影响？

4）如何进行后测？

在下面的讨论中，我们将使用虚拟研究作为案例来回答上述问题。该研究旨在比较运用电脑写作教学是否比传统的教学手段更为有效。为此，研究人员进行了一项实验研究，以比较这两个方法的效果。

1.1 选择受试者

进行实验，首先要决定实验组和对照组的人数。根据统计要求，每组人数不得低于30，比较组的数量不低于两个。从哪儿选择这些受试者？是随机从所有的学生中选择，还是从同一层次的学生中选择？是否选择自然班？尽管随机抽样是控制干扰变量的最佳方式，但采用这种方法可能会扰乱正常的教学秩序，同时，随机抽取的学生不可能在同一时间上课。因此，只好通过匹配一些重要的特征，来选择自然班级。对照组和实验组应该在写作能力、二语总体水平和专业等方面差不多。参加该项研究的有两个或多个班级（见表7.3）。

对照组	实验组
一个文科班（30人）	一个文科班（30人）
	一个文科班（30人）（可以不参加）

表7.3：对照组和实验组的班级

由于理科学生比较熟悉电脑，可能理科学生和文科学生从中的得益不一样。这样，还可以增加两个理科班，其中一个是实验组，另一个是对照组。

如果实验在暑假进行，研究人员能有更多的自由。一方面暑假教学没有必要按照国家统一大纲进行教学，另一方面上课的时间可以不受学校规定课表的限制。

1.2 前测

前测的目的是弄清楚两组学生在实验前的差异。因此，前测应涵盖可能造成实验效果混乱的主要差异。就本案例来说，研究人员至少要测试学生总的英语水平，其中包括英语写作能力。最好使用权威性的测

试，以保证考试的效度和信度。如果实验对象是大学生，可以使用非英语专业的大学英语四级、六级考试和英语专业的英语四级、八级考试。

前测的条件应该相同。如果一个班在上午考，另一个班在下午考；一个班在第一节课考，另一个班在最后一节课考，这些差异都可能影响学生的考试成绩。

前测试卷如何批改也是一个影响实验研究内部效度的重要因素。对于只有惟一正确答案的选择题来说，评分的公正性毋庸置疑。然而，给作文评分牵涉到主观判断，为了增加评分的信度，最好采取以下 3 个措施。

1）每篇作文至少由两位评分人独立评分。

2）实验班和对照班的作文应该混在一起，以避免评分人对某班的偏见。

3）给评分人统一的评分标准。

如果采用随机抽样分配实验组与对照组，可以不用前测。但有条件进行前测，可以进一步检查两个比较组是否有差异。如果没有明显的差异，后测的差异则具有更高的信度。

1.3 实验处理

实施实验处理时，有许多事情必须考虑。我们还是以前面提到的机助二语写作教学和传统二语写作教学的比较研究为例，研究人员必须考虑：怎么给实验组上课？怎么给对照组上课？实验组和对照组完成同样多的家庭作业吗？是同一个老师还是不同的老师来教？是不是在相同的自然条件（如作息时间表和环境）下授课？学习内容的差异、家庭作业数量上的差异、上课环境的不同以及教师之间的差异都可能影响因变量的变化。如果没有充分注意这些差异，最终的结果就会很难解释。两组之间的差异究竟是由计算机辅助教学引起的呢，还是由其他干扰变量引起的？

当研究人员对干扰变量做出控制决定时，问题一般比想像的要复杂得多。例如：如果要同一个老师教两个组，明显的好处是教师这一变量得到了控制。然而，不知道这位老师对两种不同的教学方法是否有偏向。即使从理论上说，她没有任何偏向，也很难保证她使用这两种方法的娴熟程度完全相同。考虑到这些问题，研究者可能决定用两个教师教，一个教师教一个班。但是，很难找到在各方面都相同的教师。最实

际的办法是选择两位大家公认的好教师，他们对在实验中所用的教学方法都很熟练。

有些实验能够立即看出效果，例如比较单词记忆方法的实验，但有些实验需要一定的时间后才能显示出效果，例如比较二语写作教学方法优劣的实验。因此决定实验处理持续的时间长短，也是实施实验处理之前必须要考虑的重要问题。

1.4 后测

在所有的实验研究中，前测可有可无，但后测必不可少。如果没有前测，设计后测时相对说来比较容易一些，因为研究人员不需要考虑测试效应。如何来实施后测呢？我们仍旧以比较两种二语写作教学方法的实验为例。假如在前测时，受试者写的作文题目是“中国应鼓励私人买小汽车吗?”，后测时，是使用相同的题目，还是不同的题目？使用相同的题目可以避免作文的难度变化。问题是有些学生记忆好，可能会把上次写过的作文再写一遍。如果使用不同的题目，一定要保证难度与前测相当。

毫无疑问，评分程序也应与前测相同。最理想的是，前测评分人与后测评分人相同。这样可增加两次测试的可比性。如果有可能，前测和后测一起评分，这样可增加两次测试评分的一致性。

2. 实施实验

实验计划越细致，实验过程中所碰到的问题就越少。然而，不管实验计划看起来多么完善，随着情况的变化，不可避免地要出现一些意想不到的问题。再者，任何一个计划都不可能完美无缺。因此，在实施过程中，需要不断进行修改。

选择实验对象的困难经常与样本的大小有关，或与实验对象分组有关。例如：研究人员可能找到两个自然班，但他们的人数不等。一个班20人，另一个班30人。如果把其中一个班换掉，相配的老师又会出现问题。遇到这种情况，最好不要换班级，因为找两个相仿的老师不容易。如果有可能，尽量把实验组放在大班。数据分析时，可以使用随机抽样的方法使两个班的人数相等。

前测必须在实际教学开始前进行，其次，两个班同时参加前测。这样可以避免因学生之间泄题导致测试无效。问题是实际的课程表常常与

研究的需要不相符。这样，必须重新安排时间。这种时间的变动必须要通知到每个受试者，否则部分学生缺考就会影响整个实验的进行，尤其是准实验研究，缺考造成的后果更加严重。有些学生对参加测试不够认真严肃。要避免这个问题，最好把前后测的成绩作为某门功课总成绩的一部分。

一旦实验处理开始，研究人员要密切注意实验的进展情况。如果有可能，研究人员应该到班上听课并做现场记录。通过实际观察，研究人员会更加了解实验处理过程中所发生的一切问题。一旦发现实验处理与原来的实验计划有偏差，或者出现了一些意想不到的问题，研究人员必须及时提出解决的办法。

关于后测，研究人员必须要处理与前测同样的问题。首先，实验组和对照组应同时参加考试；第二，应防止学生缺考；第三，后测与前测所进行的环境要相同。只有这样，才能减少干扰变量的影响。

小结

在实验研究中，研究人员在控制干扰变量的同时，对一个或多个自变量进行调控，测量自变量对因变量的影响，旨在确立因果关系。研究者对自变量的调控称为实验处理。实验处理涉及至少两组受试者或两种不同情况的比较，以检验假设的因果关系。社会科学实验研究中的因果关系与我们日常理解的因果关系有些不同，它强调的是概率倾向。实验中有两种效度：内部效度和外部效度。这两种效度对实验都非常重要。所谓内部效度是指所调查的自变量在多大程度上能够解释因果关系。外部效度是指研究结果的推广性，即这一研究结果在多大程度上能够应用到其他研究对象和其他环境中去。影响内部效度的因素是各种各样的：有些因素与环境有关；有些因素与测量有关；有些因素与选择实验对象有关；有些因素则与实验处理有关。外部效度也受到一些因素的影响，这些因素有的与环境有关，有的与选择实验对象有关。为了提高实验的效度，可以采取各种措施来控制干扰变量的影响，其中包括人为控制和统计控制。实验研究就研究人员的控制程度可分为前实验研究、准实验研究和真实验研究。真实验研究有两个特点：有对照组和随机抽样；准实验研究有对照组，但没有随机抽样；前实验研究所受的控制最少，既没有对照组，也没有随机抽样。为了实施既定的实验研究，研究人员需

要仔细考虑研究对象的数量，研究对象的选择方法，前后测所涉及的各种要素以及实验处理的方式。

讨论题

1. 问卷调查与实验研究之间有哪些差别？
2. 在什么情况下，可以说两个或多个变量之间是因果关系？
3. 哪些因素影响实验研究的效度？
4. 从国际学术期刊上找一篇实验研究的论文，就其内部效度和外部效度做一些评价。
5. 使用虚拟例子来说明3种不同的实验研究。
6. 在从事实验研究时，应考虑哪些问题？

第八章 个案研究

个案研究属于定性研究。它所涉及的研究对象人数不多，但考察的内容维度多，且具有相当深度。本章首先解释什么是个案研究，接着描述个案研究对象的选择和数据的收集。

第一节 什么是个案研究

个案研究在应用语言学领域内应用广泛，特别是外语教学与外语习得研究。什么是个案研究？简单的定义是：通过多渠道收集数据，全面、深入地研究一个或几个个案。下面是个案研究的例子。

例 1：

有个 6 岁的男孩，学英语单词的方式特别有趣。与其他小孩不同，他通过图画来记忆生词。Chen（2000）把这个男孩作为个案研究的对象，对他的英语学习观察了 6 个月，重点研究他如何记忆新单词。

例 2：

Dong（1998）研究了网上课程的案例，即：南京大学的学生与美国拉道夫—麦肯女子学院的学生共同阅读赛珍珠的一些作品，然后在网上对他们所阅读的作品进行交流。Dong 观察他们上课，并把他们交流的电子邮件和作业收集起来，以分析这门课程的利与弊。

例 3：

为了对问卷调查的结果进行补充说明，Wen（1993）对两个学生进行了个案研究。这两个学生高考入学的语文和英语的成绩几乎一样，她们两人都想把英语学好，但入学两年后，她们在全国英语专业四级统考中，英语成绩出现了惊人的差异。Wen 通过访谈，观察和记日记等方式收集了她们学习英语情况的大量数据。经过对数据的深入分析，终于发现了她们英语专业四级统测中出现成绩明显差异的原因。

例 4：

Wang（2000）挑选了不同英语水平的 16 名英语专业学生（4 个年级，每个年级 4 名）参加这项研究，要求他们每人写两篇作文，并用有声思维（think-aloud）的方式把他们的写作过程记录下来。她的研究问题是：随着外语水平的提高，在外语写作中，母语的使用是减少还是增加？

以上的例子说明个案研究可能只牵涉到一个案例，如例 1 和例 2；也可能牵涉到两个或多个案例，如例 3 和例 4。例 1 和例 2 称作为单案例研究。例 3 和例 4 称作为多案例研究。多案例研究除了对每个案例进行深入分析外，还要比较不同案例，找出它们之间的异同点。

到底什么是个案呢？任何一个现象都可称作为个案，因此很难给出一个简短明确的答案（Miles & Huberman，1994；Theodorson & Theodorson，1969）。个案可以是一个外语学习者，也可以是一位外语教师；可以是一个班，也可以是一所学校；可以是一个决定，如：语言政策、外语课程安排、教学方法和外语教学科目；也可以是某个人的特征，如：词汇学习策略、动机等。

根据 Stake（1988）的看法，个案研究应该有一个明确界定的研究范围，其界定的范围一方面要体现整体性和完整性，另一方面也要突出研究的重点。从这个定义出发，我们可以看出个案研究的几个重要特征。第一，个案研究重点明确，不需要研究所有细节；第二，对重点方面的研究应该在一定的背景下进行；第三，必须把重点方面看成是某个系统的一部分，而不是一个孤立的因素。

第二节　选择研究对象

与问卷调查和实验研究不一样，个案研究的对象不需要随机抽样。个案研究对象的数量一般比较少，即便采用随机抽样，意义也不大。然而，不需要随机抽样并不意味着个案的选择可以随心所欲。下面将介绍如何选择个案研究对象。

1. 是单个案还是多个案？

个案研究时是选择单个案例还是多个案例？这取决于研究的问题。在医药界，一个得了罕见疾病的病人往往被列为个案进行研究，其研究

结果值得发表。同样，在应用语言学领域，如果某个外语学习者或外语教师有些独特的地方，就值得进行个案研究。然而，要完成硕士论文或博士论文，一般需要选择多个案或对单个个案进行长期跟踪研究。

2. 选择个案研究对象时应考虑的因素

在选择个案研究对象时，首先要考虑如何选择。选择的成功取决于多种因素，其中最主要的有两个因素：研究对象的特征和研究对象的态度。

首先研究人员要确定这些个案哪些特征相同，哪些特征不同。除了要明确研究的特征外，还要明确哪些特征必须加以控制。以前面提到的例3为例。Wen的研究目的是弄清楚为什么有些学生入学考试分数几乎相同，两年后在全国英语四级考试中，成绩会出现很大差异。经初步分析认定学习策略可能是造成差异的主要原因。为了进一步了解学生使用策略的情况，Wen有意选择了两个同班同学，并且她们同住一个寝室。此外，她们在家庭背景、入学时的语文水平、入学时的英语水平以及学习动机方面都差不多。但她们的努力程度却与四级考试成绩相背（见表9.1）。

从表9.1可以看到，除了四级考试成绩和努力以外，Wen所选择的两个个案在其他方面几乎没有差异。这样做，就有可能把她们在学习策略上的差异看作造成四级考试成绩差异最重要的原因，而排除其他因素可能引起的争议。否则即使在使用策略上存在着差异，我们也有充足的理由说：策略不一定是影响成绩的最重要因素。Wen有意选择了高分而课外花的时间少，低分却课外花的时间多的人。她的用意非常清楚，她想说明尽管低分者比高分者付出的时间多，但策略上的差异最终仍旧导致学习成绩不理想。

编号	性别	年龄	父母职业	高考语文	高考英语	学习目的	努力程度	专业四级成绩
01	女	19	父亲：大学教师 母亲：大学教师	75	95	2.67	20.5	90.50
06	女	20	父亲：医生 母亲：护士	75	96	2.67	40	64.25

表9.1：个案研究对象的个人简况

第二个要考虑的因素是研究对象的态度。与问卷调查和实验研究形成鲜明对照的是，个案研究只有几个研究对象，经不起中途流失。再者，参加个案研究的人需要付出较多的时间与精力独立完成一系列的任务。从这个意义上说，参加个案研究的人一定要乐于合作，态度认真。否则，个案研究就无法进行。怎么才能确保被选择的研究对象合作、认真呢？首先，在确定候选人参加这项研究前，研究人员一定要明确地说明研究的内容与要求，一般情况下总会碰到有些学生不愿意参加，平均拒绝率为20%，差生拒绝率比好生高。如果差生和好生人数的平衡对研究极其重要，在选择候选人时就要充分考虑到这个问题。最后，参加个案研究的人必须都是志愿者。研究者要采用一定的方式对他们的合作表示谢意。最简单的方法是送给他们一些小礼物，诸如词典、图书或文具之类。此外，还可以在研究结束时，帮助他们分析学习中的问题，给他们提供一些建议，这对想提高英语水平的差生来说，可能更为有效。

第三节　收集数据

在个案研究中，研究人员常常用不同的方式从多个来源收集数据。常用的方式有访谈、有声思维和日记等。研究人员通常在一项研究中，使用两种或多种方式。下面就这些常用的方式逐一进行介绍。

1. 访谈

访谈是收集定性数据的重要形式。所谓访谈是指研究人员通过谈话向访谈对象提出一系列的问题，以了解访谈对象在某个方面的情况。正如前面所提到的，问卷调查有开放式的问题和封闭式的问题。访谈也有开放式的问题和封闭式的问题。下面我们仅就访谈中开放式的问题展开深入的讨论。

1.1　访谈的利与弊

访谈的优势表现在多方面。通过访谈，可以获得一些用其他方法不易或不能获得的信息。例如：使用其他方法，你难以观察到一个人的感情、思想和用意；你无法观察到以前发生的事；你也无法观察到你不允许到场时发生的情况。但是，访谈中你可以面对面地询问这些情况。访

谈可以使我们了解他人看问题的视角。我们对访谈的假设是：别人的看法有意义，可以为人所知。然而，我们对从访谈中得出的数据仍要谨慎对待，这是因为在访谈过程中，访谈对象可能因记不清或有偏见，导致对有些问题的描述不准确。此外，访谈对象只能报告在意识范围之内注意到的事情。最后，如果研究人员缺乏访谈技巧，也可能大大降低访谈数据的效度和信度。

1.2 访谈的种类

根据访谈者在访谈中的自由程度，访谈可分为 3 种：非结构化访谈、半结构化访谈和结构化访谈。

非结构化访谈又称作开放式访谈。这种访谈给访谈者许多自由。访谈时就像自由聊天，甚至被访者没有意识到自己在接受访谈。开放式访谈的好处是有人情味，没有拘束，容易谈得深入。例如：老师可能分别找学生谈话，讨论期中考试中出现的问题，其谈话的真正意图是想了解学生自我评价能力和自我分析问题的能力。

半结构化访谈，也称作半开放式访谈。访谈者根据事先准备好的问题与被访者进行交谈，但访谈不一定按照原来问题的顺序，也不一定按照原来的措辞。访谈者可以有所发挥，见机行事，碰到有用的信息要紧追不放，所提问题层层递进。这要求访谈者拥有高超的访谈技巧：离题深追但不扯得太远，果断返回主题而又不显得生硬。

结构化访谈由一整套开放式的、事先规定好了的具体题目组成，访谈者按照同一顺序，统一措辞向被访者提问。访谈者不能随意追加问题，或改变问题的顺序。控制访谈问题数和统一提问的程序可以提高访谈的信度。这种访谈特别适合被访者人数众多，访谈者有好几个的情况。其缺点是灵活性低、过于机械，缺乏人情味。

3 种访谈的共同特点是访谈对象用自己的语言表达各自的观点。尽管在措辞和问题的顺序上有些变化，但答题的原则不变：答案应该是开放式的。也就是说，访谈者在访谈中不给被访者提供答题所用的词语和类别。

访谈的分类看起来泾渭分明。然而，在实际的访谈中，访谈者常常根据其目的和临场情况，将 3 种类型的访谈交替使用。

1.3 准备访谈计划

对于半结构化访谈和结构化访谈来说，应该准备好访谈计划。访谈

者必须决定问哪些问题，这些问题的顺序是什么，需要多少细节，访谈多长时间，以及这些问题的措辞，等等。具体地说，准备访谈计划需要做3件事：1）决定想获得信息的类别；2）确定问题的顺序；3）选择问题的措辞。

1.3.1 信息的类别

和问卷一样，访谈一般可以获取5类信息：1）经历或行为；2）感情；3）观点；4）知识或能力；5）背景情况。这5类问题可以与不同的时间段相结合：过去、现在和未来。一旦决定要收集哪一方面的信息，就可以确定访谈的具体问题了。

1.3.2 访谈问题的顺序

非结构化访谈比较灵活，访谈问题很少有固定的顺序。然而，结构化访谈必须有一个固定的顺序。关于访谈问题的顺序，这里有3点建议：1）先问容易的问题；2）先问此时此地的问题；3）先问有趣的问题。

1.3.3 访谈问题的措辞

提问的目的是要被访者提供所需的信息，因此访谈问题的措辞直接影响访谈对象的答题内容。提问是一门艺术。在定性研究中，符合要求的访谈问题至少应该是开放式的。真正开放式的问题可以让访谈对象自由发挥，不设任何限制。我们知道，在回答封闭式问题的时候，研究对象只能从事先预设的有限的答案中选择一个。换句话说，在封闭式的问题中，各种答案已提供。许多访谈者认为设计开放式问题，就是不提供结构化的答案。当然事情并没有那么简单。有些问题表面上看没有结构化答案，但实际上它是把预设的答案隐藏在问题之中了。

请看下面的问题："你对在语言实验室里上口语课满意程度有多高？"从表面上看，这是一个开放式的问题。然而，仔细一看会发现问题的答案已有了预设，即：你满意的程度，如："相当满意"、"满意"、"基本满意"等。实际上，问题的答案已被访谈问题的措辞限制死了。因此，如果把刚才的问题改成"你认为在语言实验室上口语课怎么样？"，那就成了一个真正的开放式问题。

真正的开放式问题，不会让被访者仅从某个特定的视角出发描述自己的情感或看法。应该允许被访者自由表述，不受任何限制。例如：

你在语言实验室内上口语课感受如何？

你对在语言实验室内上口语课的看法如何？

此外，真正的开放式问题不能设计为是非题。访谈的目的是要求被访者深入、详细地描述自己的经历、感情、观点和知识。是非题不仅没有鼓励访谈对象开口，反而造成被访者处于两难境地，因为访谈对象不知道他们是否只需要提供一个简单答案。下面是 Patton（1990：297－298）举出的例子。

（年轻人约会回来。）

甲：你知道你回来晚了吗？

乙：知道了。

甲：玩得开心吗？

乙：开心。

甲：电影不错吧？

乙：不错，很好。

甲：这么说，很值得一看了？

乙：当然。

甲：你认为我会喜欢这部电影吗？

乙：不知道。也许会吧。

甲：关于晚上情况，有什么想告诉我吗？

乙：没有。我想就这些。

从上述例子可以看出，是非题把访谈变成询问或考问，而不是交互式的谈话。

1.4 访谈策略

为了获得高质量的数据，在访谈中可以使用录音、沉默式追问、重复式追问、肯定式追问这样一些访谈策略。

1.4.1 使用录音机

在访谈中，不要过分相信和依赖自己的记忆力。如果被访者不反对，所有的访谈都需要进行全程录音。磁带不仅可以长时间保存原始信息，而且可供其他研究人员检验或使用。准备录音机时，最好准备一台备用。录音磁带的质量要保证，因为日后将录音转写成文字时，磁带需要多次反复审听，质量差的磁带经不起这样反复使用。

每次访谈前，必须检查一下录音机。如果录音机使用电池，备用电池的数量要充足。好的录音机有电池显示器，当电池用得差不多时，就及时换新电池，免得影响录音质量。如果打算用交流电，也要准备一些

电池，以防停电。

1.4.2 追问

成功访谈的关键是学会有效地追加问题，获得更多有价值的信息，也就是说，访谈者无需过多参与交际，但能从被访者那儿获得很多信息。访谈中，有多种追问方式，下面介绍 Bernard（1994）推荐的几种。

最容易的，可能也是最难的一种追问方式是“沉默式追问”。沉默式追问就是保持沉默，等候访谈对象继续说。沉默往往伴随着点头或“嗯”声以示同意。沉默式追问有时比直接追问获得的信息还要多。有些访谈对象比较健谈，基本上不需要追问就能说出大量的信息。而有些访谈对象喜欢思考，说话不慌不忙。经验不足的访谈者一旦发现访谈对象沉默往往马上就用语言去追问，其实，访谈对象可能在思考，或内心在总结，或准备说出重要内容。如果用语言去追问，就可能打断他/她的思绪。必须指出，沉默式追问风险比较高，新手应避免过多使用。设想被访者真的无话可说，而访谈者不仅不加以引导，相反沉默不语，这肯定让被访者尴尬不已。要想学会有效地使用沉默式追问，必须要在访谈实践中用心揣摩、反复练习。

另一种追问叫“重复式追问”，也就是说，重复被访者所说的最后一句话，然后要求其继续说下去。当被访者在描述某一过程或事件时，重复追问这一方式尤其有用。例如：你要被访者描述其阅读过程。一旦访谈对象停止描述，你就可以说：“噢，我明白了，你刚才说你先看一下题目，然后猜测文章的大意。那你接下去做什么呢?”这种追问是中性的，不需要重新指明访谈的方向。但如果重复式追问用得过多，你可能会听到访谈对象生气地说：“你为什么老是重复我说过的话?”

第三种追问方式是肯定式追问。这就是说，访谈者通过发出一些肯定的声音如：“嗯，是啊。”“噢，我明白了。”“对的。”等，以鼓励被访者继续说下去。

1.4.3 日常交谈式风格

所谓日常交谈式风格，就是指访谈者与被访者之间的交际就像日常谈话一样，有问有答，问答相连。特别需要注意的是，访谈者后来所提的问题必须与被访者前面的回答有联系。这一点不容易做到。一方面，访谈者事先准备了一整套问题，必须使访谈不能走题；另一方面，访谈者的提问不要使被访者感到在被询问或审问。这里需要有灵活应变和即席处理话题的能力。

1.5 访谈中常见的问题

通过观察某些研究生的模拟访谈实习，我们发现有以下几个常见的问题。第一个问题是许多学生不能忍受短暂沉默。一旦被访者不讲话，他们就急于向被访者发问，这样容易打断被访者的思考。第二个问题是他们经常问是非题，被访者只能给予简短的肯定或否定回答，接下去他们就不知道该问什么。第三个问题是当被访者离题时，他们不知道如何把话题重新拉回来。最后一个问题是访谈者说得太多，基本上把答案都告诉了被访者。

培养有效访谈技巧需要大量的实践与反思。所谓反思是指对每一次模拟访谈进行自我评价，找出不足之处，下次加以克服。经过多次实践与反思，就能娴熟地掌握访谈技巧。

2. 有声思维

自 20 世纪 80 年代以来，随着认知研究兴起，有声思维作为收集数据的一种方式，逐渐取得了合法的地位。许多研究人员如 Cohen（1987）、Goodman（1989）、Guo & Wen（1998）、Flower etal（1990）、Hosenfeld（1984）、Hudelson（1989）、Lauer & Asher（1988）、Lu（1997）、Raimes（1985）以及 Zamel（1983，1987）都曾采用这一方式对学生的二语写作和二语阅读进行研究。这些研究的结果使我们对学生的认知过程有了更加深刻的认识。下面将回答以下 3 个问题：1）什么是有声思维？2）如何进行有声思维？3）有声思维有哪些局限性？

2.1 什么是有声思维？

3－7 岁的小孩在玩游戏时，常常会自言自语，说出心里的想法。有声思维就如同小孩自言自语一样，要用语言说出内心活动的内容。在外语研究中，有声思维要求研究对象讲出内心的思维活动。与儿童自然的有声思维形成对照，外语学习者的有声思维必须经过培训学习才能有效进行，换句话说，对于学习外语的成人来说，有声思维不再是一个自然行为。如果没有适当的训练，成人肯定难以按照要求完成有声思维的任务。

下面是两个有声思维的例子。

例 1：Guo 的研究（1997）

Guo 用有声思维研究学生在看图作文时使用母语的程度。20 名高二

和高三的中学生参加了该项研究。经老师推荐，这些学生性格外向，比较健谈，适于完成有声思维的任务。

他们接受了为期两周的有声思维培训。Guo 先以读古文示范。他一边读古文，一边大声把心里所想的讲出来。接着，要求受试者用有声思维完成两件事：1）阅读一篇古文，用有声思维说出他们的理解；2）看一幅缺损的图画，用有声思维描述出他们理解不完整图画的过程。

接受培训以后，受试者要根据所给的图画，写一篇 100 词的英语作文。在写作过程中，必须一边写，一边把大脑里所想的讲出来。进行有声思维时使用的语言不限，既可以是英语也可以是汉语，但有声思维一定要自然，怎么想就怎么说。总共有 8 幅图画，描写的是一个学生在上学的途中，为了救一个盲人的生命，结果上学迟到了。Guo 对学生写作时的有声思维进行了全程录音，同时把学生的作文也全部收集起来。

Guo 把所有的磁带录音整理成书面文字。然后比较在有声思维中使用母语和使用外语数量上的差异，接着通过定性分析，鉴别使用母语的不同功能，最后提出外语写作中母语使用模式的假设。

例 2：Lu 的研究（1997）

Lu 想研究外语水平与外语阅读中使用学习策略之间的关系。4 名南京大学外语专业的学生参加了这项研究。其中两名是日语专业的，两名是俄语专业的。一名日语专业的学生和一名俄语专业的学生组成初级组，他们只学了一年英语，另外两名学生组成中级组，他们学了 7 年英语。这 4 名学生在自己所学的专业中都属于高分组。

Lu 选择了两篇适合受试者水平的文章，并要求他们在阅读文章时尽可能多地报告他们理解文章的过程。他们的口头报告被全程录音。Lu 坐在旁边静静观察。只有当学生沉默超过 1 分钟时，Lu 才要求受试者继续讲话。阅读没有时间限制。

为了保证受试者熟悉有声思维的操作程序，他们试读了一篇文章。由于他们是大学生，很快就掌握了有声思维。

Lu 把口头报告的录音整理成书面文字。以 Block（1986）的策略分类为基础，适当做了些修改，对文字材料进行了分析处理。下面是本次研究的结果。

外语水平与使用理解策略之间的关系非常复杂。初级组学生受外语水平的限制不能使用某些高水平的理解策略，或者使用成效不明显。此外，初级组学生使用的阅读策略较少，使用的频率较低，使用的质量较

差。同时他们在处理生词和难句子时，通常信心不足，表现得焦虑不安，惧怕困难。

2.2 如何进行有声思维？

有声思维听上去容易，但做起来难。要成功使用有声思维，至少要满足4个条件。第一，受试者愿意合作，即按照研究人员的要求来进行有声思维；第二，受试者接受足够量的培训；第三，磁带的质量必须保证；第四，受试者进行有声思维时，研究人员的表现要训练有素，能够积极促成有声思维的顺利完成。

2.2.1 选择理想的研究对象

有声思维对于成年人来说，不再是自然的语言行为。有些人平时就不善于言谈，还有些人尽管健谈，但对有声思维态度不积极。这些人即便接受了培训也不能按照要求完成有声思维的任务。因此从这个意义上说，所谓“理想的”的研究对象，就是那些能在经过培训后，顺利进行有声思维的人。为了保证有声思维的质量，最终的受试者要经过3轮选拔。第一阶段，选择研究对象参加培训。选择标准为性格外向，善于言谈。第二阶段，要选择那些能够自在进行有声思维并且愿意合作的人。根据经验，第二阶段的淘汰率约为40%，因为他们对有声思维感到不舒服。第3次选择是在数据收集以后进行。那些在有声思维中有多次长时间沉默的不适合参加最后阶段的数据分析。这一阶段的淘汰率大约为10%到20%。由于最终能够参于数据分析的人要经过3次筛选，因此人选参加培训的人员一定要有足够的余量。

2.2.2 进行有效的培训

在对受试者进行培训时，正常的程序是研究人员先进行示范，并对有声思维的操作进行详细说明，然后，至少给研究对象两次机会练习。受试者的第一次练习可在语言实验室进行，大家同时做，既可以节省时间，又可减少紧张情绪，同时各人的有声思维都可以录在磁带上，供研究人员发现问题。第二次练习的时候，最好单个进行。这样，如果他们对有声思维有误解，研究人员可以及时帮助解决。通过个别观察，更容易挑选出“理想的”受试者。

在培训过程中，一个潜在的问题是研究人员的示范可能会起误导作用。这就是说研究人员的有声思维可能给受试者提供了某种暗示。例如：你想研究在外语写作中使用母语的程度。示范时，你一会儿用英

语，一会儿用汉语把你所想的都讲出来。研究对象可能认为他们也应该这样进行有声思维。因此，在该项研究中，选择涉及两种语言交替使用的示范内容，就不恰当。最好选择一个与语言学习无关的话题，如：猜汉字所缺的部分，或解谜。因此示范内容的选择必须谨慎小心。

2.2.3 准备高质量的录音机

访谈的问题相对独立，彼此之间不一定有逻辑联系。如果录音机出了毛病，访谈可随时停下来。但有声思维不同，一旦录音机出了故障，受试者的思维中途被迫打断，下次很难恢复。因此，用来进行有声思维的录音机应该仔细地检查，确保万无一失。

此外，最好每个受试者都使用新电池，以防在有声思维过程中电量不足或突然用完。最后，要确保录音机的录音开关处在工作状态。

2.2.4 研究人员的行为专业化

当研究对象进行有声思维时，研究人员该怎么办？一般说来，研究人员的行为要符合专业人员的要求。所谓行为专业化是指研究人员所做的一切都要训练有素、符合专业规范，最终有助于研究任务的完成。此时此刻，研究人员最好坐在受试者的视野以外，静静地观察受试者的表现。如有必要，可以记笔记，但不要走动或发出声音，不应干扰研究对象的思维活动。

如果受试者在有声思维的过程中出现稍长时间的沉默怎么办？

要不要问“你在想什么？”最好的办法是发出一种事先约定的信号，提醒受试者停止沉默。

2.3 有声思维的局限性

最近20年以来，有声思维逐渐获得了外语学习研究者的认同。越来越多的研究者使用这一方式，深入了解外语学习者的思维过程。然而，这一方法有一些先天不足。我们在解释研究结果时，一定要考虑到这些不足。

如前所述，有声思维只适用于少数最终能按照既定计划成功完成任务的研究对象。从这个意义上说，有声思维只适用于部分受试者。

因此研究人员必须要承认用这种方法得到的研究结果不一定能够应用到其他人身上。

有声思维毕竟不是正常的成人行为，因此不管培训多么有效，当要求成人进行有声思维时，他们的思维过程都可能在某个方面不那么真实

自然。因此，我们不清楚由有声思维而获得数据在多大程度上能反映思维的真实情况。当然培训越有效，所收集的数据就越真实。

有声思维要求研究对象如实讲出他们的内心活动，但不可避免的是，有些心理活动不可能讲出来，这是因为思维比讲话速度快得多。此外，讲话呈线性式，思维呈多维式。讲话只能一句一句地讲，而思维可以同时在不同层面展开。因此心口不一不足为怪。从这个意义上说，有声思维不可能百分之百报告其内心活动。为了弥补其不足，研究人员可进行后续访谈以获得更多的信息。

尽管我们意识到有声思维的局限性，但这种方法仍是目前了解学习者内心思维的最佳方法。随着认知研究的发展，人们对外语学习的心理活动越来越感兴趣。到目前为此，我们还没有科学仪器设备能了解人类的内心思维。有声思维至少在某种程度上可以满足我们的研究需要。

3. 日记

日记是我们研究外语教学的一个重要工具。它可以记录外语认知过程或其他外语教学过程中学习者和教师的感情、态度、行为、反思和对认知过程的认识。

3.1 日记的界定

根据 Bailey（1990：215，Nunan 转引 1994：120）的定义，研究中使用的日记是“第一人称对语言学习或阅读经验所做的叙述，通过不间断的、真实的记录，以分析出反复出现的模式和特别的事件”（p.215）。这一定义强调了对日记的 3 点要求。第一，必须是研究对象自己记日记；第二，记日记必须要有一定的频次，其内容必须真实；第三，日记应记录与语言学习有关的情况，以便研究学习者在语言学习方面的异同点。

Bailey 的定义没有区别日常生活的日记与用于研究的日记。第一，在日常生活中的日记是个人感情、观点、随感的记录，一般不允许别人阅读，因此它可以是片言只语，可以不合乎逻辑，也可以不合乎语法。然而，一旦日记用作研究，它就失去了个人隐私的特点。记日记的人要用浅显明白的语言，便于研究者理解。第二，日常生活中的日记内容随心所欲，不受任何限制，而研究用的日记必须包含与研究目的相关的情况。第三，日常生活中的日记主要记录刚刚发生的事情，而研究用的日记既可以记录此时此地发生的事情，也可以记录彼时彼地的行为和想

法。后者我们称为反思日记。

根据以上的差别，笔者把研究用的日记定义为“一个人用明白晓畅的语言对自己外语教和学经历的如实的书面记录，这些记录都必须与具体的研究问题相关”。

3.2 实例说明

一些外语研究人员把他们自己外语学习的经历记录下来，以便更好地了解外语学习的过程。例如：Rivers（1983）在其日记里把她如何学习第六种语言记录下来了。Bailey（1983）在日记中记录了自己学习法语的丰富经历。

很多研究人员也要求研究对象记日记，以便他们研究这些个案的外语学习过程。下面我们来看一下他们是怎样通过日记来收集数据的。

例 1：Wen 的研究

在 Wen（1993）的外语学习者可控因素研究中，从 242 个研究对象中选择了 10 名进行个案研究。作为个案研究的一部分，要求这 10 名学生记一周的日记，以了解他们课外学习活动的组织、安排以及他们在使用学习策略方面的差异。下面是 Wen 发给参加个案研究对象的注意事项：

请记一周的日记。日记中要记录每天课外所进行的英语学习活动。日记的内容一定要包括以下几个方面：

1）你进行了什么活动？为什么要进行该项活动？

2）你进行每项活动所花费的时间是多少？

3）你从事每项活动的身体和心理状态怎样？（是否精力旺盛？是否精力集中？）

4）你认为这项活动怎么样？（是否对提高你的语言水平非常重要？）

5）你在从事这项活动时，采用了哪些策略？（请尽可能详细些。）

由于本次记日记的目的不是检查他们的语言水平，因此对他们记日记的语言没有要求。只要能把意思表达清楚，他们可以用英语，也可以用汉语，还可以英汉语混用。

依据这 10 个学生的问卷和访谈的数据分析，结果显示高分组和低分组使用策略都很主动积极。然而，日记却显示了高分组与低分组在定性方面的差异，这种差异在问卷和访谈的数据中没有显示。其原因是高分组往往比低分组确定的要求要高，因此当高分组没有达到要求或完成

任务时，他们常常责怪自己。而低分组对自己的要求不高，往往过高地估计了自己。例如，通过日记，Wen 发现低分组往往比高分组在使用管理策略方面差。下面就是低分组同学的两篇日记：

1991 年 10 月 23 日 星期三

下课后我去了图书馆。我坐在“语言参考书”那一角。这里有许多参考书。我经常选择不同的书阅读。今天我对《英语习语》特别感兴趣。里面的习惯用语是按字母顺序排列的，大部分都很有用，例如：“Nothing ventured, nothing gained”……我抄下了一些我喜欢的句子。我一边抄，一边记，花了一个下午，抄了 70 多个句子，才抄到字母“c”。我经常做类似的事情：不管它有没有用，只要我喜欢它，我就宁愿花时间。我就是用这种态度来进行英语学习的。尽管今天下午看起来学得不多，但我相信“Many a little makes a mickle.”随着时间的推移，我的英语会有进步的。（第 10 个研究对象）

1991 年 10 月 26 日 星期六

今天下午我看了萧伯纳的 Pygmalion。全书有 300 多页，前面是作者生平及其作品简介。一个下午我花了 3 个小时，只读了简介的一部分，共 10 页。我一边读，一边记下一些好句子和有关作者的一些简况。遇到难懂的句子，我用蓝色的圆珠笔做上记号。碰到生词我就查词典，并把词的解释抄下来……通过阅读我学到了很多东西。我认为作为英语专业的学生应该多读点名著。

很明显，上述两次学习活动都带有很大的偶然性。她事先没有明确的目标，因此对阅读材料和阅读方法都没有进行有意识的选择。首先，《英语习语》是工具书，不适合作为课外阅读材料。她企图通过阅读与抄录来记住那些习惯用语，显然效果不会理想。抄写的过程往往比较被动，没有自己的积极思维活动，抄一遍，大脑中不可能留下很深的痕迹。即使有自己的积极思维，也不可能一次记住彼此没有意义联系的 70 多条习语。因此，在很大程度上可以说，她是浪费了一个下午的时间。

课外阅读萧伯纳的 Pygmalion 似乎没有什么可以值得责备的地方，但不妥当的是，她采用精读的方法来阅读这样的作品，既耗时费力，又效果不佳。遗憾的是，她并没有发现上述活动收效甚微，相反还误认为这是学习英语的好方法。

例 2：Schmidt 和 Frota 的研究

Schmidt 和 Frota（1985）要求一个英语本族语者记录自己在巴西学习葡萄牙语的经历。日记表明他学习的经历可以分为 3 个阶段：（1）与当地人自然交谈；（2）通过上课学习；（3）通过自然交际伴随事后的自

我反思。下面3段日记摘要正是记录了他在3个不同阶段学习的认知与情感经验。

日记摘要1：第一周日记

我痛恨自己不能够与周围人交谈。我成天喜欢讲话，不喜欢沉默。碰到巴西人很容易，但语言是唯一的障碍。我注意到在巴西街上不管性别、年龄或阶层，没有人介意向别人要香烟的行为。昨天晚上我碰到个漂亮的、有身份的年轻女子，有男朋友陪同。她拦住我，向我要支烟。如果我带一盒烟到海边，肯定不到一个小时所有的香烟就会被要完。这就是说，我可以碰到20个人。今天P和我在海边，有个家伙上来要烟，并坐下来想跟我说话。他问我是否是美国人。我说"sim"。他又说了一些东西，我不明白，因此无法回答。然后他说："看来与你交谈有困难。"我虽然能听懂他说的意思，但我记不得他所用的词语。（第242页）

日记摘要2：第四周日记

昨天P和我参加了一个葡萄牙语班。班上共有11个来自不同国家的学生。上课的老师年纪轻，但课上得很好。她用葡萄牙语向我们作自我介绍。她说，我是X，我的名字叫X。…… 我也是英语老师。我是单身，没有结过婚，没有孩子。我有一个应用语言学的学位，等等。她在教室里转来转去，问我们类似的问题：你叫什么名字？你从哪里来？你做什么工作？你有孩子吗？我们中的大部分能够回答其中的一些问题，例如我知道我在大学的职称。每个人都很快地从别人的答案中学到新东西。其余的同学轮流互相介绍，直到全部介绍完毕。快下课时，她把SER的词形变化和一些新单词写在黑板上。好极了。这比born bia和沉默好多了。我相信在我离开这里之前，这样的问题我会被问成百上千次。因此昨天晚上我出去与4个人说了话。讲得不错。我被邀请参加第二天晚上的聚会。当然，我很快就没有话可说了，也很快听不懂别人对我说的话。这使得我迫切希望回到课堂上去学习。（第243页）

日记摘要3：第11周日记

H和我在外面吃晚餐。他不停地在抱怨自己的工作。我试图说：看上去你的工作并没有使你不舒服。从他脸上的表情看，他不明白我的意思。我查了词典，发现comfortable在葡萄牙语里应该是comfortavel。但我脑中突然闪过这个词只能描述椅子之类的物体，而不能描述人。几分钟以后，H用nao deve.说了一些事情。我曾经学过DEVER的意思是"have to"或者是"must"。我想nao deve +动词的意思应该是"don't have to"，deve nao +动词的意思是"must not"。但从H说的话来看，很清楚是"should not"。因此我学到了一些东西。总的说来，H是个缺乏交谈艺术的人。在我说的话中，其他人都能理解的，他偏偏听不懂。当我不明白他的时候，他所能做的就是重复。（第246页）

从以上的例子我们可以看出用于研究的日记与日常的日记之间的区别。用于研究的日记最终能够向我们揭示外语学习过程中的某些本质性特征。

3.3 日记的利与弊

使用日记比有声思维和访谈更加方便。首先，就研究人员来说不需要像有声思维和访谈那样花费许多时间，因为有声思维和访谈均需一个一个地进行，并要把录音一一转写成书面材料。第二，在技术要求上没有有声思维和访谈那么高。有声思维对研究对象的要求非常挑剔，在收集数据前还要通过一定的培训。关于访谈，除了精心设计访谈的问题，访谈数据的质量还取决于研究人员的访谈技巧。而采用日记收集数据，只要满足一个条件：即向研究对象讲清楚记日记的要求。否则，等到日记收集起来，才发现日记内容不能为回答研究问题服务，就为时太晚。

尽管日记可以给我们提供用其他方法很难收集到的数据，但日记所收集的数据也面临许多挑战和批评。首先，人们对其效度表示怀疑，也就是说，人们无法肯定日记所记的情况是否反映实际情况。很明显，研究对象是在事情发生后描写他们的学习经历和学习态度，有些描写可能会不太准确，尽管不是有意的。第二，日记通常记录研究对象所能意识到的东西。最后，日记有时写得不清楚或很难懂。

3.4 使用日记中常见的问题

使用日记收集数据简单易行，操作方便，但记日记过程中是否存在问题，一般中途无法得知。只有等日记全部交上来的时候，研究人员才能判断交上来的日记是否符合研究的需要。而与之相对的有声思维和访谈，研究人员可以在操作过程中发现问题，及时纠正。因此，研究人员必须预测研究对象可能会出现哪些问题，给他们以明确的要求，并尽量避免这些问题的发生。

研究生常遇到的一个问题是在研究对象的日记中，许多信息与研究问题不相关，而相关的信息又不足量。造成这一问题的原因是研究人员没有向研究对象解释清楚要他们写什么。第二个问题是，如果交待记日记要求时带有偏向，日记所收集的数据就会含有这一偏向。例如：在一项日记研究中，所有研究对象都说，他们过多使用连词的原因是教师教学所致。当我们读到这样的日记时，肯定要问为什么所有的研究对象都过度使用连词呢？原来研究人员对日记的要求中就提到，要他们解释为什么过多使用连词，由此可见，研究人员的要求本身就带有偏向。

小结

个案研究牵涉到的研究对象人数不多，收集的数据属于定性数据。根据研究目的，我们可以进行单个案研究和多个案研究。在选择研究对象时，必须注意两个问题：1）研究对象的哪些特征需要控制，哪些特征需要有差异？2）研究对象愿不愿意合作？收集个案研究的数据有3种方法：访谈、有声思维和日记。访谈是研究人员与访谈对象之间的交流，通过交流，研究人员可以了解到访谈对象的情况。这种方法可了解到其他收集数据方法所收集不到的情况，如：访谈对象的情感、他们对过去、现在和未来的一些看法。其最明显的缺陷是访谈对象自己报告的数据仅限于有意识的层面，可能会被有意或无意歪曲。按照研究人员的自由程度划分，访谈可以有3种：非结构化、半结构化和结构化访谈。对于半结构化访谈和结构化访谈，应该准备好访谈的问题，这些问题必须是真正开放式的问题。为了有效地进行访谈，最好使用以下几个策略：使用录音机、层层追问和采用谈话的方式。有声思维是研究对象在进行语言行为操作时，把心里所想的讲出来。使用这一方式耗时、费力。被选为有声思维的研究对象经过培训后，要能够按照要求完成规定的任务。有效的培训是有声思维成功的关键。日记也是个案研究中收集数据的一个重要工具，它可以记录在外语教学中发生的情况或对以前发生的事情进行反思。使用日记比有声思维和访谈更为简便、容易操作，但有时日记所报告的数据不一定能完全反映现实。

讨论题

1. 什么是个案研究？请举例说明。
2. 非结构化访谈和半结构化访谈的区别是什么？
3. 访谈有哪些利和弊？
4. 用录音机把对一个学生的访谈录下来，然后把录音放给其他学生听，让大家评论访谈中存在的问题。
5. 如何运用有声思维来收集数据？如何确保所收集数据的高质量？
6. 在什么情况下有声思维最有效？为什么？
7. 有声思维有哪些利和弊？
8. 什么时候使用日记作为收集数据的方式？请举例说明。
9. 日记作为收集数据的方式有哪些利和弊？
10. 日记作为研究工具有哪些常见的问题？

第九章　微变化研究

微变化研究（microgenetic analysis）是最近20年来认知心理学领域中兴起的一种研究方法。该方法主要用于探究认知发展的轨迹与机制，侧重研究群体或个体发展过程中的变异性（variability）。追溯其历史，早在70年前Vygotsky和Werner就提出了这一研究方法的基本构想（Miller & Coyle，1999），但直到20世纪80年代才有人付诸实践。迄今，已有多种国际权威杂志刊登了与这一研究方法相关的研究论文。

目前虽然微变化研究还没有成为主流研究法，但从它的发展态势和所具有的优点可以断定，该方法必定会得到越来越广泛的运用，最终会与横向研究法（cross-sectional design）和纵向研究法（longitudinal design）互为补充，成为认知科学中不可或缺的研究方法。

第一节　什么是微变化研究？

微变化研究要求研究者在某个相对短的时间段内，高频度观察、记录受试者学习和掌握某个观念、某个知识点，或某项微技能的过程，并对数据进行定量、定性的精细分析，以揭示群体内部与个体本身所表现出的变异性。

1. 微变化研究的基本要素

Siegler & Crowley（1991）指出，微变化研究必须具备3个基本要素：1）观察的时间周期要恰当；2）观察频度要高；3）数据分析要精细。

观察的时间周期可以是几十分钟，几小时，几天，几个星期或几个月。例如Siegler和Crowley（1991）的观察周期长达11个星期，每个星期3次；Schauble（1996）的观察周期为6个半小时，也有的观察周期只

有 30 分钟（Bjorklund et al，1992；Miller & Aloise-Young，1996）。这里最关键的是观察周期要覆盖变化的全过程。最为理想的安排是：观察开始于变化即将产生之前，结束于被观察行为的变化已经停止，进入了稳定期。

Siegler 和 Crowley 运用飓风袭击某个城市为例来说明观察频度。设想我们拍摄了两张照片：一张在飓风之前，一张在飓风之后。人们可以通过比较这两张照片，评估出这个城市在飓风中所遭受的损失，但据此无法推断遭受损失的过程。要想了解全过程，最理想的是拍电影。如果没有条件拍电影，还有一种退而求其次的做法：连续拍摄一系列静态照片。只要照片的密度恰当，人们就可以从中推测出飓风袭击城市的全过程。微变化研究与拍摄系列静态照片的原理相似。研究者高频度全程记录某种行为的系列变化，就有可能从中发现变化的关键时刻、变化的速度、变化的原因以及变化的广度和深度。这里的关键是观察频度。如果频度不高，受试者行为变化的采样密度不足，就难以推断出复杂的变化过程。

微变化研究的第三个要素是精细分析受试者在量和质两个方面的变化情况。以研究某个策略的微变化为例，如果数据分析只局限于变化速度与认知准确性的数量变化，而没有相关的定性分析，那就失去了微变化研究应有的价值。引起速度与准确性变化的原因比较复杂。有的可能源于新策略的使用，有的可能源于使用已有策略频率上的改变，也有的可能源于使用已有策略熟练程度的提高。精细分析策略使用的情况就能够区分不同原因所起的不同作用，以及不同原因在何种情况下起作用。正是有了量和质两方面的精细分析，微变化研究才能弥补传统研究认知变化的不足。

2. 微变化研究的优势与局限性

Miller & Coyle（1999）总结了微变化研究的 4 大优势。第一，该方法能够直接观察到正在发生中的变化。目前许多研究者采用了横向研究法，偶尔也有人用纵向的跟踪研究法考察儿童思维的变化。尽管这些研究提供了不同年龄段儿童思维能力的宝贵数据，但无法说明儿童的思维能力是怎样从一个年龄段发展到另一个年龄段的。这些数据只是某一个发展阶段上的几个静态点。由于两个静态点之间的间隔太大，变化的过程与机制就难以推断。而微变化研究在这方面具有明显的优势。

第二个优势是，微变化研究可以考察变化的不同方面。Siegler

(1995）提出了研究变化的5个维度：路径（path)、速率（rate)、广度（breadth)、变异性（variability）和根源（sources of change)。根据Siegler的理论，变化路径指的是受试者所进行的系列尝试。速率指的是变化的快慢。从群体数据中看到的发展往往是比较缓慢的渐变过程，而事实上对于不同的个体、不同的时间段，变化有可能非常急速。因此群体数据一般掩盖了变化速率这一问题。广度指的是受试者对概念或者技能应用的范围。变异性指的是在前3个维度上的个体差异。根源指的是变化的原因。这5个维度可能随着不同的任务、不同的情境、不同的个体而形成差异。微变化研究最大的贡献就是它能识别出最可能发生变化的条件、最可能的原因，以及最可能使原因起作用的方法，进而能够解释变化的复杂性。

第三个优势是，这个方法能够发现行为中的变异性。横向研究设计与群体数据分析只能让研究者注意某个年龄段儿童的典型行为。例如某个年龄段的儿童具有某种心理结构、信息处理极限、规则和策略。相反，微变化研究考虑儿童个体在不同的时间所进行的不同尝试。而群体表现的规律不仅掩盖了组内差异，而且掩盖了个体行为本身的不稳定性。

第四个优势是，这个方法很灵活。它能用于研究不同的概念、不同的技能、不同的知识点。此外，这个方法也适合不同的理论流派，例如皮亚杰派、新皮亚杰派、维果斯基派、信息处理流派都愿意运用这个方法研究认知发展。

尽管微变化研究有许多优势，但是它也有其明显的局限性。第一，该方法对受试者须逐个观察、逐点记录、耗时费力，因此受试者的数量有限。第二，这种方法只适用于动机强的受试者，因为整个研究的成败取决于受试者能否全程认真参加多个实验，并且在每个试验中，从头到尾都能积极进行尝试。第三，尽管观察密度高，数据很丰富，但在有些情况下，仍旧难以分清受试者已具备的能力和变化后的能力。根据以上的局限性，我们可以断定，该方法的推广有相当的难度。

3. 微变化研究成果在理论层面上的贡献

根据Miller和Coyle 1999年的估计，微变化研究的项目大约只有40到50个，其主要原因是：该方法费时耗力，操作困难。尽管研究项目数量还不够多，但其研究结果已对现存的认知发展理论提出了有力挑战。例如传统的认知发展研究者，一般强调某个阶段的典型且有规律的

表现，而忽视了变异性。例如皮亚杰提出人的认知发展全过程可以分为以下四个阶段：1）从出生到2岁为感知运动阶段；2）从2岁到6至7岁为前操作思维阶段；3）6至7岁到11至12岁为具体操作阶段；4）11至12岁到成人为形式操作阶段。Siegler（2002）用阶梯模型来比喻皮亚杰的发展论（见图9.1），并指出这些递进式的发展描述只是反映了群体发展的宏观趋势，而不能反映儿童认知发展的真实情况。事实上，在发展过程中，不仅群体内部存在差异，而且个体本身也存在差异，变异贯穿于发展的全过程，忽视差异性，也就谈不上真正意义上的发展研究。

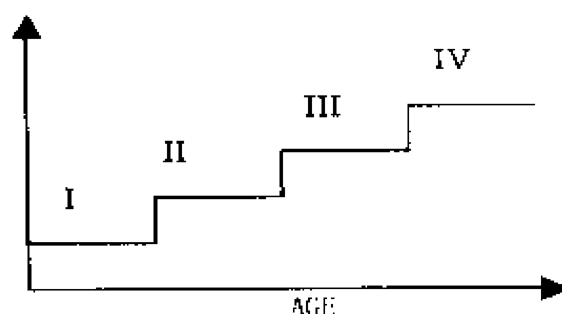


图9.1：阶梯模型
（Siegler，2002：32）

与传统阶梯理论针锋相对，Siegler（1996，2002）根据微变化研究的系列成果，提出了波浪重叠型理论（overlapping waves theory）。该理论强调发展的不均衡性和变异性，并指出发展时快时慢、时有时无、时进时退。例如某个策略的成功，受试者不一定记住它；某个策略的失败，受试者也不一定放弃它。儿童有时暂时偏离目标，为的是以后能够更好地接近它。策略的变化既表现在频率上，又表现在有效性上。在熟练运用某个策略的过程中，定量与定性的变化总是互相交织在一起。

第二节　认知心理学领域中微变化研究的案例

下面我们将运用 Siegler & Stern（1998）所进行的一项研究“Conscious and Unconscious Strategies Discoveries：A Microgenetic Analysis”来说明微变化研究的具体操作与价值。该项研究的目的是要揭示儿童在解决算术问题中发现捷径策略（shortcut strategy）的意识程度。所谓捷径策略

就是指儿童在计算（a+b-b）这样的加减逆运算题目时，不经过计算，就能够一眼看出该题目的答案为a。也就是说他们知道+b-b互相抵消，实际是不加不减。不知道捷径策略的儿童就需要花费比较长的时间，先用加法得到a+b的和，再从a+b的和中减去b。Siegler和Stern希望通过系列测试来了解原来不掌握捷径策略的儿童是怎样学会使用这样的策略，并考察这个学习的过程是个渐进的过程，还是个突变过程。此外，他们还要分析儿童在这方面的表现是否存在着明显的差异性。

1. 研究设计

1.1 参加试验人员

该项研究的受试者为31名德国小学二年级学生。经过前测筛选，他们均为不会运用捷径策略进行逆运算的儿童。根据实验要求，他们被随机分成普通加减混合题组（mixed group）和加减逆运算题组（blocked group）（后面简称混合题组和逆运算题组）。

1.2 实验过程

实验在放学后的活动中心里进行。整个实验包括8次测试，分别由3位女老师负责。所有受试儿童每星期在课外活动中心参加1次测试，连续参加8次。测试题写在卡片上。每张卡片上只有1道题，受试者1次只能看到1道题，完成每道题后，需要立即回答老师提出的问题：你是怎样得到这个结果的？老师一边用马表准确记录学生回答每道题所花的时间，一边记录答案，此后还要记录学生对解题策略的描述。8次测试题的结构形式见表9.1，每次测试的题目内容不一样，题目的顺序随机编排，但题目的难易度保持不变。

	全部受试者	
测试1（前测）	40道测试题，一半为加减逆运算题，形式为a+b-b；一半为普通加减混合运算题，形式为a+b-c。	
	混合题组	逆运算题组
测试2，3，6（练习）	每次有20道测试题，一半为逆运算题，一半为普通加减混合运算题。	每次只有20道逆运算题，没有普通题。

（见下页）

（接上页）

	全部受试者
测试5，7（练习）	每次有20道测试题，一半为逆运算题，一半为普通题。
测试8（后测）	共有48道测试题，6种形式，每种形式有8道题。这6种形式包括：a+b-b，a-b+b，a+b-a，a-b-b，a+b+b，a-b+a.

表9.1：8次测试题的结构形式

1.3　数据分析

数据分析分5步进行。第一步分析受试者完成每道题的时间，时间值呈双峰分布：39%的计算为1-4秒；8%为5-7秒；53%为8秒或8秒以上。

第二步归纳受试者对解题结果的解释。解释共有以下3种：

1）全程计算：先做加法24+36=60，然后再做减法60-36=24

2）部分计算：先做了加法24+36=60，但没有做减法，就一口报出答案24。

3）无计算：没有经过运算，一口报出答案24。

第三步依据解题时间、受试者的解释，以及是否有外显的计算行为，将策略由低到高分为5种水平（见表9.2）。

		解题时间	受试者的自我解释	外显行为
低级 ↓ 高级	1．全程计算策略	4秒以上	描述了加减的全过程	能观察到计算的过程
	2．部分计算策略	4秒以上	只描述了加法的过程，没有减法的步骤	能够观察到部分的计算过程
	3．无意识捷径策略	4秒或少于4秒	解释不清楚计算过程	不能观察到计算的过程
	4．计算+捷径策略	4秒以上	描述了捷径策略的使用	不能观察到计算的过程
	5．捷径策略	4秒或少于4秒	描述了捷径策略的使用	不能观察到计算的过程

表9.2：计算策略的界定

第四步从策略变化的路径，变化的速率，变化的原因，变化的广度和变化的个体差异5个方面多层次分析了受试者8次测试的数据。

2. 研究结果

与传统研究方法所得的结果相比，该项研究结果*特别值得强调的可以归纳为以下3个方面：1）多层次分析策略的变化；2）区分无意识与有意识策略的使用；3）认识新旧策略关系的复杂性。

2.1 多层次分析策略的变化

该方法在3个层次上分析了策略的变化：1）逐次分析8次测试策略的使用频率；2）逐人、逐次归纳使用5个不同策略水平的顺序；3）逐人、逐次分析受试者对策略使用的尝试，特别是有意识和无意识使用捷径策略的前后。与传统的数据分析结果相比，该项研究既有实验组与对比组宏观、整体行为的比较，也有组内个别行为差异的比较；既有两组使用不同水平策略的一般顺序，还有每个学生捷径策略使用前后行为的精确分析。

2.2 区分无意识和有意识策略的使用

确定儿童运用策略的意识程度一直是心理学家难以解决的问题。该项研究将受试者的解题时间作为策略使用的间接依据、受试者随后对解题过程的陈述为策略使用的直接依据。研究者将两种依据结合，逐次、逐题判断受试者使用策略的意识程度。实践证明，这种方法行之有效，且操作性强。运用这个方法，该项研究发现绝大部分的受试者在有意识使用捷径策略之前就已经进行了多次无意识的尝试。这一结果对我们认识儿童习得策略的复杂过程提供了重要实证依据。

2.3 认识新旧策略关系的复杂性

儿童计算逆运算题的策略可以根据水平不同，由高到低分为5种，捷径策略为最高水平。该项研究表明，儿童使用策略有个循序渐进的过程，但这个过程不是一个简单的线性模式。较高水平的新策略往往会在较低水平策略使用的过程中偶尔出现，此时儿童一般不能对较高水平策略的使用进行语言表述。即使等到他们能用语言描述时，较低水平的策

* 由于篇幅所限，这里不能报告该项研究的详细结果，有兴趣者可以阅读 Siegler & Stern (1998) 的英文研究报告。

略仍旧与新策略的使用共存。例如逆运算组到第2次练习时，已有81%的儿童发现了捷径策略，但在之后的5次练习中，大约有三分之一的习题计算仍旧使用了低水平的全程计算策略。

以上的微变化研究结果为我们了解儿童发现捷径策略的过程提供了多层面、多纬度的立体描述。传统的研究结果只注重整体的行为变化趋势，而忽略了变化的复杂性和多样性。

第三节　微变化研究在二语习得领域内的应用前景

1. 二语习得领域中微变化研究的意义

从20世纪60年代开始，二语习得作为应用语言学的一个相对独立的分支展开研究，80年代进入快速发展时期（Ellis 1997）。有的研究中介语中存在的错误，有的研究个体差异对二语习得的影响，有的研究学习者内部因素与二语习得的关系；有的研究宏观因素（社会、文化、经济、政治因素）在二语习得中的作用。简单说来，研究的焦点可以归为两个：1）研究结果（product），即外语学习者的中介语体系本身；2）研究过程（process），即外语学习者中介语发展的过程。但就其实质来看，绝大部分研究仍旧是描述二语学习的结果，而不是从发展的角度探究中介语变化的过程与机制。

Ali Shehadeh（2002）曾经指出，在众多对Swain（1985）提出的输出假设研究中，大部分研究都是横向设计，研究的焦点都是调整过的输出本身，而不是这样的输出能否成为学习者中介语发展的动力。从这个意义上说，现存的研究焦点是学习者的语言表现，而不是他们的语言习得。绝大部分的实证研究局限于考察学习者个体差异和情境因素对输出的影响，而不是考察调整过的输出能否帮助二语学习，以及怎样帮助二语学习。例如有不少研究分析了性别差异对二语学习者输出的影响。他们发现有证据说明男女在交谈中表现不同。男学习者多说，而女学习者多听。由此他们推断男性的二语学习多得益于输出，女性多得益于输入。事实上，这样的研究根本没有触及输出是怎样影响二语习得这一议题。在分析了众多所谓的二语习得实证研究后，Ali Shehadeh呼吁研究者们要真正研究二语习得的过程与原因，而不是二语学习者的语言行为。

我们认为微变化研究为二语习得研究者提供了一个新手段，或者说一个新视角。如果我们采用微变化研究，精心设计操作任务，让受试者反复多次进行尝试，研究者详细记录全过程，并对所得的数据从变化的路径、速率、广度、原因和变异性5个不同维度，进行精细的定量与定性分析，我们就有可能发现变化的轨迹与机制。同时这样的研究还可以检验英国心理学家 Nick Ellis 在 2002 年 6 月年度特刊《第二语言习得研究》上发表的主题论文"Frequency effect in language processing"中提出的"频率是语言习得关键"假说（文秋芳，2003）。微变化研究的重要特点之一是要高频度观察受试者的反复多次尝试，就其实质，我们也是在考察频率对二语习得的作用。

2. 二语习得领域中值得微变化研究的课题

二语习得领域中微变化研究至少可以从两方面入手：1）纠错研究；2）微型语言项目习得研究。就语音纠错而言，我们选择的研究对象要明确、具体，一次研究选择一个重点。例如某些大学生在个别英语语音的发音上存在困难，有的分不清/r/、/l/和/n/；有的分不清/s/和/θ/。一次只研究某一个困难，否则观察难以进行。我们可以选择几个有相同困难的学生进行高频度的、有针对性的练习，每天一次，每次 20 分钟。练习内容精心设计，练习项目包括相关的单音、单词、短语、句子和短文。对语法纠错研究，我们可以选择中国人常犯的过去时错误或第三人称 he、she 不分的错误。让受试者进行高频度练习，例如每星期 3 次，每次 20 分钟，持续 5 个星期。每次练习的内容有单句，有对话，有独白等多种形式。无论是语音还是语法纠错研究，每个受试者练习必须单个进行，每次练习之后需要进行访谈。研究者一定要逐人、逐次、逐点进行记录，并进行全程录音或录像。这样的研究可以提供丰富的定量和定性数据。从定量的角度，我们可以知道学生出错次数的变化，出错广度与深度的变化，练习次数的作用；从定性的角度，我们可以了解学生出错的困难所在，错误率下降的原因。

与纠错研究不同，微型语言项目习得研究必须是全新的语言现象，受试者中没有人学过。例如我们可以选择儿童，或者中学生。设想我们计划研究受试者学习定语从句的过程。研究开始前，我们可以通过测试，挑选出从未学过定语从句的受试者。然后让受试者经过反复、多次的尝试，观察周期可以定为两三个星期，每个星期 3 次，每次 20 分钟。

研究者必须精心设计每次的练习内容。记录、分析数据的方法与纠错研究相同。

3. 应用语言学领域中微变化研究尝试案例

Zhu*（2003）运用微变化研究考察了频率（frequency）在英语学习中的作用。该研究探讨了频率对阅读理解和词汇学习的影响。

3.1　研究问题

该研究要回答以下3个问题：

1）输入频数如何影响二语学习者的阅读理解？

2）输入频数如何影响二语学习者的词汇学习？

3）输入频数是否对阅读理解和词汇学习产生同等的影响？

3.2　受试对象

受试对象为4名非英语专业四年级的大学生，其中1男3女，年龄在22岁到25岁之间，他们的TOEFL和GRE词汇成绩分别列在表9.3中。他们都非常乐意参与本项实验，并表示会认真完成每次实验的任务。

研究对象	TOEFL成绩	GRE词汇成绩
W	647	630
X	637	670
Y	643	650
Z	653	700

表9.3：4名研究对象的TOEFL和GRE词汇成绩

3.3　研究工具

3.3.1 阅读材料

阅读材料是一篇中等难度的英文说明文，题目为“美国与第一次世界大战”。全文共有472个单词，附有25个生词的边注，其中10个词需要受试者掌握。课文内容没有超过受试者的知识范畴。

3.3.2 阅读与词汇测试

阅读测试包括与阅读材料相关的7道多项选择题，6次测试的内容

* 这里仅报告了Zhu微变化研究的部分结果。

不变。词汇测试包括前测和后测。前测包括文章中25个加边注的单词和25个干扰词，目的是挑选出4名受试者都不熟悉的10个词作为后测的目标词。在前测中凡是熟悉的英文单词受试者就写出中文的对应词，不熟悉的就跳过去。后测共有6次，每次测试包括15个生词，其中10个是目标词，5个是干扰词。干扰词每次有变化，10个目标词不变，但出现的先后次序有变化。受试者根据Paribakht & Wesche（1997）制定的词汇熟悉程度5级量表，标注自己对每个单词的熟悉程度。

3.3.3 实验步骤

实验前两天，所有受试者参加词汇摸底测试（前测）。前测结果表明4名受试者在25个阅读材料中出现的25个单词中有12个生词，其中两个属于科技专业词汇，不适合用于后测。

所有受试者参加了6次实验，连续进行了3天，每天上下午各一次。每次实验按照下列程序进行：1）阅读所给材料10分钟，材料中没有测试题；2）回答阅读理解题7分钟，回答时不允许参考阅读材料；3）运用15分钟完成词汇测试，不允许参考阅读材料或词典。

6次实验完成后的一个星期，又突然要求受试者参加了一次只含有10个目标词的词汇测试。测试方法与前面相同。

3.3.4 数据分析

对单词熟悉程度的等级		可能得分
I 不记得见过这个词。		0
II 记得见过，但不知道词义。		1
III 能够给出中文意思或英文同义词。		2
IV 能够在句子中正确表达词义。		3
V 能够用词造句，不仅词义正确，而且语法形式正确。		4

表9.4：单词熟悉等级程度量表

（Paribakht & Wesche，1997：181）

阅读测试满分为7分，答对1道题给1分。词汇测试中只给10个目标词评分。按照5个熟悉程度等级，分别给以0分，1分，2分，3分和4分。满分为40分。具体给分方法见表9.4。

根据表9.4，除了第一、第二个等级的得分不和别的等级有交叉外，其他3个等级都有可能降级得分。换句话说，如果受试者自我报告对单词的熟悉程度达到第三级，但给出的词义不对，这时只能按照第二等级给分。

3.4 研究结果

3.4.1 输入频数对阅读理解测试成绩的影响

表9.5－9.8分别列出了4名受试者6次阅读理解的测试成绩。从4名受试者6次的总分来看，前3次的成绩总体上呈直线上升趋势，但把6次的成绩放在一起，情况就变得复杂多了。受试者Z在第3次测试中，成绩达到了满分，其后没有再出现变化。受试者Y在第3次测试中答对6道题以后，就再也没有进步了。受试者X在第3、4、5次得到6分，但第6次又回到了5分。受试者W虽然在第3次和第4次测试中得到满分，但在第5次和第6次测试中又回到了6分。

根据上述结果，我们可以推断，输入频数对阅读理解有一定的帮助，3次是理想次数，或者说3次是关键次数，在3次之内，每增加1次，阅读理解的成绩就有所提高，如果少于3次，个人潜力就可能得不到充分的发挥，但超过3次以后，输入频数的增加对阅读成绩的正面影响不明显。

实验次数	第1题	第2题	第3题	第4题	第5题	第6题	第7题	总分
1	1	0	0	0	1	1	1	4
2	1	1	0	1	1	1	1	6
3	1	1	1	1	1	1	1	7
4	1	1	1	1	1	1	1	7
5	1	1	0	1	1	1	1	6
6	1	1	0	1	1	1	1	6

表9.5：受试者W 6次阅读理解测试成绩

实验次数	第1题	第2题	第3题	第4题	第5题	第6题	第7题	总分
1	0	1	0	1	0	1	0	3
2	1	1	1	1	0	1	0	5
3	1	1	1	1	1	0	1	6
4	1	1	1	1	1	1	0	6
5	1	1	1	1	1	0	1	6
6	1	1	0	1	1	0	1	5

表9.6：受试者X 6次阅读理解测试成绩

实验次数	第1题	第2题	第3题	第4题	第5题	第6题	第7题	总分
1	0	1	1	0	1	1	0	4
2	1	1	0	0	1	0	0	3
3	1	1	1	1	1	1	0	6
4	1	1	1	1	1	1	0	6
5	1	1	1	1	1	1	0	6
6	1	1	1	1	1	1	0	6

表9.7：受试者Y 6次阅读理解测试成绩

实验次数	第1题	第2题	第3题	第4题	第5题	第6题	第7题	总分
1	1	1	0	0	1	1	1	5
2	1	1	1	1	1	0	1	6
3	1	1	1	1	1	1	1	7
4	1	1	1	1	1	1	1	7
5	1	1	1	1	1	1	1	7
6	1	1	1	1	1	1	1	7

表9.8：受试者Z 6次阅读理解测试成绩

3.4.2 输入频数对词汇测试成绩的影响

表9.9－9.12分别列出了4名受试者6次词汇测试成绩。从4名受试者6次的总分来看，随着输入频数的增加，词汇测试成绩总体上是直线上升趋势。第4次好像是个进步快慢的分界线。Z自第4次开始，一直得满分40分；W自第4次开始，一直得39分；总起来说，他们在前

4次词汇测试中，进步比较快，后面两次的进步明显放缓。X、Y第6次的测试成绩都是35分，这就是说这两个人尽管在3天中有6次学习新词的机会，但他们仍旧有一半的单词没有达到会用的地步。这似乎意味着一个新词从初学到会用的程度，有人需要4次学习的机会，有的需要6次以上的机会。此外，各人学习新词进步的速度不均衡。受试者W、X和Z在第2次词汇测试中成绩提高最快，他们分别提高了8分、6分、6分，而受试者Y在第4次中提高最快，一次提高了9分。有趣的是，一个星期以后的词汇测试成绩与一个星期以前第6次词汇测试的成绩几乎没有差别，这似乎表明前面6次学习词汇的效果得到了很好的保持。

实验次数	单词										总分
	tonic (n.)	by all odds (adv.)	emphatically (adv.)	veritable (adj.)	hidebound (adj.)	agrarian (adj.)	steep (v.)	instigate (v.)	suffuse (v.)	engulf (v.)	
1	3	2	2	3	2	4	3	4	0	3	26
2	3	3	4	2	3	4	4	4	3	4	34
3	4	3	4	3	3	4	4	4	4	4	37
4	4	4	4	3	4	4	4	4	4	4	39
5	4	4	4	3	4	4	4	4	4	4	39
6	4	4	4	3	4	4	4	4	4	4	39
后测成绩	***4***	***4***	***4***	***3***	***4***	***4***	***4***	***4***	***4***	***4***	***39***

表9.9：受试者W 6次词汇测试成绩

实验次数	单词										总分
	tonic (n.)	by all odds (adv.)	emphatically (adv.)	veritable (adj.)	hidebound (adj.)	agrarian (adj.)	steep (v.)	instigate (v.)	suffuse (v.)	engulf (v.)	
1	1	2	3	3	3	2	2	1	2	1	20
2	2	2	3	3	3	2	2	4	3	2	26
3	2	2	3	3	4	2	3	4	3	3	29
4	3	3	3	3	4	3	4	4	3	3	33
5	4	4	3	3	4	3	3	4	3	3	34
6	3	4	3	3	4	4	3	4	4	4	36
后测成绩	***3***	***4***	***4***	***3***	***4***	***3***	***3***	***4***	***4***	***3***	***35***

表9.10：受试者X 6次词汇测试成绩

实验次数	单词										总分
	tonic (n.)	by all odds (adv.)	emphatically (adv.)	veritable (adj.)	hidebound (adj.)	agrarian (adj.)	steep (v.)	instigate (v.)	suffuse (v.)	engulf (v.)	
1	1	1	2	1	1	2	1	3	1	1	14
2	1	2	2	2	1	2	1	3	3	2	19
3	1	2	3	2	2	3	2	3	3	2	23
4	3	3	3	3	3	4	3	3	4	3	32
5	3	3	3	4	3	3	3	3	4	4	33
6	3	4	3	4	3	3	3	4	4	4	35
后测成绩	***3***	***4***	***3***	***4***	***3***	***4***	***3***	***3***	***4***	***4***	***35***

表 9.11：受试者 Y 6 次词汇测试成绩

实验次数	单词										总分
	tonic (n.)	by all odds (adv.)	emphatically (adv.)	veritable (adj.)	hidebound (adj.)	agrarian (adj.)	steep (v.)	instigate (v.)	suffuse (v.)	engulf (v.)	
1	3	3	3	2	2	4	3	3	1	4	28
2	4	3	4	3	3	4	3	3	3	4	34
3	3	4	4	3	4	4	3	4	3	4	36
4	4	4	4	4	4	4	4	4	4	4	40
5	4	4	4	4	4	4	4	4	4	4	40
6	4	4	4	4	4	4	4	4	4	4	40
后测成绩	***4***	***4***	***4***	***4***	***4***	***4***	***4***	***4***	***4***	***4***	***40***

表 9.12：受试者 Z 6 次词汇测试成绩

3.4.3 输入频数是否对阅读理解和词汇学习产生同等的影响？

从 6 次阅读理解测试和 6 次词汇测试成绩变化的情况来看，输入频数对词汇学习和课文理解都产生积极影响，但词汇学习达到最佳水平的时间稍稍滞后于阅读理解。此外，输入频数对渐进式的词汇学习作用更加明显。

3.5 对 Zhu（2003）研究的评价

朱靖的研究从两个维度检查了输入频数对阅读理解和词汇学习的影响。没有她高密度的观察及精细的分析，我们很难得知对于阅读理解和

词汇学习的最佳输入次数，我们也很难得知阅读理解和词汇学习的不同特点。当然，我们必须承认 Zhu（2003）的研究还仅仅是微变化研究的初步尝试，从严格意义上说，她研究的深度还不够。

最后必须强调的是，微变化研究的切入点一定要小，设计要周密，观察频度要高，分析要精细，这样的研究好比微雕艺术，看上去很小，但很精致，很耐看。这样的研究结果不仅能够为我们发展二语习得变异性理论提供实证依据，而且可以为二语教学的改进提出具体建议。我们建议有志于二语习得变异性研究的老师和学生可以尝试运用微变化研究的方法，开展系列试验，共同探讨中国学生中介语的发展过程与机制。

小结

微变化研究（microgenetic method）是最近 20 年来认知心理学领域中兴起的一种研究方法。该方法主要用于探究认知发展的轨迹与机制，侧重研究群体或个体发展过程中的变异性（variability）。实施该方法的基本要素包括恰当的观察周期、高频度的观察和精细的数据分析。该方法能够从路径、速率、原因、广度和变异性 5 个不同维度观察到认知观念或行为的变化。这一方法具有广泛的适用性，能够弥补传统横向研究法和纵向研究法的某些不足。但该方法耗时费力、只适用于研究动机强、愿意配合的受试者。笔者认为该方法在二语习得研究领域具有广泛的应用前景。二语习得研究者可以运用该方法分析频率作用对二语学习者学习某个微型语言项目的过程或改正某个固化错误的过程。

讨论题

1. 什么是微变化研究？
2. 微变化研究的优势是什么？不足之处是什么？
3. 从事微变化研究时应该特别注意哪些问题？

第十章　统计学入门

生活在电脑时代是幸运的。处理定量数据原本要花费大量的人力、物力，现在可以运用统计软件，转瞬之间就能得到准确、详尽的分析结果。当然，运行统计软件时还需要研究人员首先做出选择，即需要采用何种程序对数据进行处理。同时统计软件也不能代替研究人员对数据分析结果做出解释。这就是说，统计软件是功能强大的工具，可以为研究者服务，但它并不能代替人的智慧。一个研究人员如果不具备统计的基本知识，即便她拥有统计软件依然不能对数据进行正确的处理。研究人员必须掌握统计学中的一些基本知识。

第一节　统计学简介

统计学可分两大类：描述性统计和推断性统计。我们首先简要介绍什么是描述性统计和什么是推断性统计。然后，讨论两个重要的术语：参数和统计量。

1. 界定

描述性统计主要是通过对数据的整理、分类来描述一组数据的全貌，揭示一个事物的总体特征。推断性统计主要通过从样本中得到的统计量来推断总体的情况，旨在得出总体的规律性特征。使用推断性统计学的最大好处是研究人员在没有研究总体的情况下，可根据样本推断出总体的情况。

描述性统计学与推断性统计学两者之间既有区别，又有联系。一方面，描述性统计学是推断性统计学的基础，使用推断性统计之前必须首先使用描述性统计；另一方面，描述性统计只是对数据进行一般的分析和归纳，若不进一步应用推断性统计对事物进一步分析，有时会使统计

结果失去意义，达不到统计分析的目的。

2. 参数与统计量

参数又称总体参数，是指描述一个总体情况的一些统计指标，用希腊字母表示，如反映总体集中情况的统计指标称为总体平均数，用 μ 表示。反映总体分散情况的统计指标称为总体标准差，用 σ 表示，有时也用总体方差 σ^2 表示。统计量是描述一个样本的数据情况，用罗马字母表示。如反映样本集中情况的统计量称为平均数，用 $\bar{X}$ 表示。反映样本分散情况的统计量称为标准差，用 S 或 SD 表示，样本方差用 S^2 表示。

只有在使用推断性统计学时，把“参数”和“统计量”区别开来才显得有意义。对总体进行全面调查，除了需要大量的人力、物力和财力以外，还需要大量的时间，因此，对参数的计算通常不现实。在这样的情况下，研究人员从总体中取出样本，计算统计量，并由此推断出参数值。

假如你是某大学的英语老师，你想知道全校所有学生使用策略和英语水平之间的关系。大学有两个校区，一万多名学生。很显然，你没有这么多的时间，也没有这么多的财力来进行全面调查研究。你可能从中选择 400 名学生作为样本，用问卷对他们进行调查。然后，用描述性统计和推断性统计对数据进行分析。例如：你使用描述性统计了解这 400 名学生常用哪些策略，得出的结果是样本的统计量。你可以使用推断性统计学研究使用策略与英语成绩之间的关系，或者根据概率了解男生与女生在使用策略方面是否有明显差异。

描述性统计学，又称描述性分析，它只是描述所收集数据的总趋势。而推断性统计学，又称推断性分析，旨在根据样本对总体进行推断，它比描述性统计学要复杂得多。

第二节　描述性统计

如前所述，描述性统计学主要通过对实验或调查数据的整理和分类，来描述一组数据的全貌，揭示一个事物的总体特征，如：频率、集中趋势和离散程度。

1. 频率

由于频率分布可以有效地归纳和总结数据，因此整理数据最简单的方法就是描述其频率分布。频率分布既可以用统计表来表示，也可以用统计图来表示，如：柱状图、线形图和扇形图。

首先，要学会如何用统计表来报告频率分布。如果数值的分布范围有限，描述频率分布就很容易。例如：调查对象对 5 分制量表的问卷回答不可能超过五个：1、2、3、4、5，这时只要统计一下所有回答，就可以获得频率分布（见表 10.1）。

选择	频率	百分比
1（坚决不同意）	30	30%
2（不同意）	45	45%
3（不置可否）	10	10%
4（同意）	5	5%
5（完全同意）	10	10%

表 10.1：频率分布

有时数值的分布范围很广，如人们的年龄、学生考试的分数等，我们就需要把原始数据进行分组或分类。这样，频率分布所关注的不是各个分数，而是各个组或类别。如何按组或类别来描述频率分布呢？这里没有固定不变的规则可循。即使原始数据相同，不同的研究人员最后分组方法也不一定相同。下面是一组新生期中考试的分数：

58，65，84，70，90，75，86，76，80，82，

83，84，69，84，85，86，72，89，75，92

一般说来，我们在描述数据频率分布时需采用以下 3 个步骤：

1）求全距；

2）定组数；

3）定组距。

所谓全距（用符号 R 表示），是指一组数据中最大值与最小值之间的差距。求全距的计算方法很简单，就是把数据中的最大值减去数据中的最小值，得数就是该数据的全距。上述数据的全距是：

R = 92 − 58 = 34。

确定组数就是要把整组数据划分为多少个等距的区组。组数的多少要看数据的多少而定。组数不能太多，也不能太少。组数太多，很难概括；组数太少，看不出重要的差异。一般情况下，组数以 5 到 10 之间为宜。

在知道全距和组数之后，接下来就可以确定分组的组距。全距与组数的比就是组距的取值。通常，我们选择 5 或 10 作为组距，如果不行的话，也可以选择 15，20，25，50 等作为组距。

频率分布必须包括所有的数据。因此，频率分布开始的值应该等于或小于该数据的最小值，频率分布终点的值应该等于或大于该数据的最大值。

组距	频率	百分比
50—60	1	5%
60—70	2	10%
70—80	5	25%
80—90	10	50%
90—100	2	10%
总数	20	100%

表 10.2：新生成绩分 5 组的频率分布

由表 10.2，可以看出该数据的组数是 5，组距是 10，获得这样表达的数据，我们就可以毫不费力地给以合理的解释。你可以说 12 个学生考得不错，其他 8 个学生仍需努力。总的来说，整个班的成绩正常。

我们还可以用 3 个分数段来描述以上数据的频率分布，见表 10.3。

组距	频率	百分比
50—65	1	5%
65—80	7	35%
80—95	12	60%
总数	20	100%

表 10.3：新生成绩分 3 组的频率分布

有时，频率并不能说明多少问题，百分位会让人一目了然，见表10.4。

分数	频率	累积频率	百分位
58	1	1	第5
65	1	2	第10
69	1	3	第15
70	1	4	第20
72	1	5	第25
75	2	7	第35
76	1	8	第40
80	1	9	第45
82	1	10	第50
83	1	11	第55
84	3	14	第70
85	1	15	第75
86	2	17	第85
89	1	18	第90
90	1	19	第95
92	1	20	第100

表10.4：带有累积频率和百分位的频率分布

百分位代表某一个具体分数在一组分数中所处的相对位置。以下是计算百分位的公式：

$$百分位 = \frac{累积频率}{N} \times (100)$$

上述公式中，累积频率是指某个分数的频率加上小于该分数的频率。如表10.4所示，65分出现1次，因此它的频率是1，而它的累积频率是2，因为它前面还有一个58分。这样1+1=2。N代表分数的总个数。

百分位可定义为某个分数超过一组分数的百分数。例如：某个学生在托福考试中的分数的百分位是第67。这就意味着她的成绩高于67%的

学生，低于 33%的学生，或者说 67%的学生考得没有她好。为了确保能真正理解百分位的定义，表 10.5 再提供几个百分位，以及它们的解释。

百分位	解释
第 95	该分数比 95%的分数要高
第 90	该分数比 90%的分数要高
第 75	该分数比 75%的分数要高
第 60	该分数比 60%的分数要高
第 50	该分数比 50%的分数要高

表 10.5：百分位及其解释

除了用表格来说明频率分布，还可以用柱状图、线性图和扇形图等统计图形来说明频率分布。见图 10.1、图 10.2 和图 10.3。

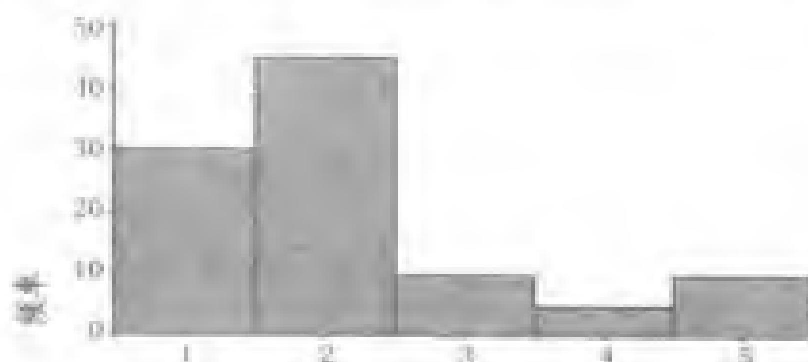


图 10.1：柱状图

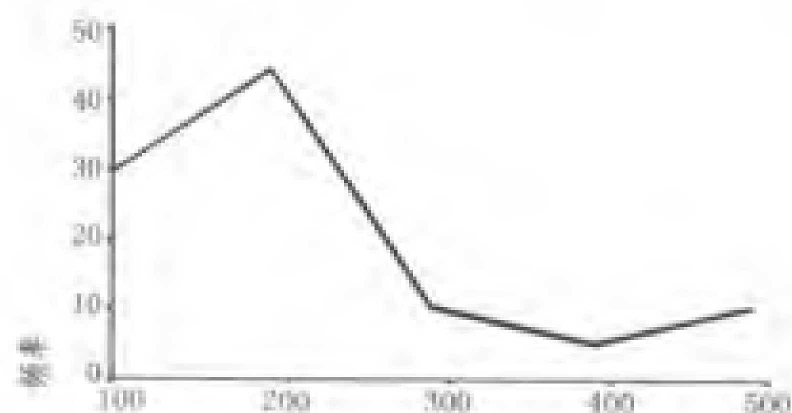


图 10.2：线性图

2. 集中趋势

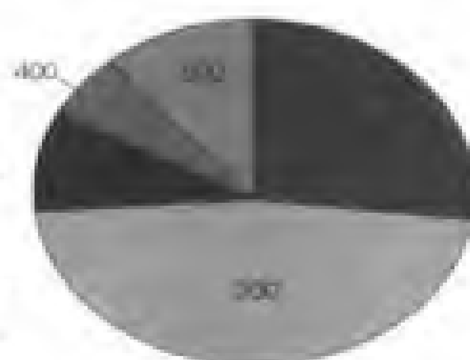


图 10.3：饼状图

一般说来，最常见的描述集中趋势的有众数（mode）、中数（median）和平均数（mean）。

众数是指一组数据中出现次数最多的数值。例如：在下列数据 58，60，65，70，80，84，84，84，84，90，92 中，众数为 84。

有时，如果在数据中有两个数值出现的次数一样多，就有两个众数。例如：在下列数据 58，60，65，65，65，65，84，84，84，84，90 中，众数是 65 和 84。

由于不需要计算，使用众数描述集中趋势最容易。然而，它有明显的缺陷，即对偶然的数字非常敏感。我们来看下面的例子：在下面的分数 58，64，65，65，65，85，85，86，88，90 中，众数是 65。但是，如果有一个考了 65 的学生碰巧考了 64，而另一个考了 86 的学生碰巧考了 85，那么，众数就由 65 变成了 85。这当中相差 20。因此，我们要认真考虑这一局限性。

顾名思义，中数是指位于一组数据中间位置上的那个数。要想计算中数，首先要把数据按照从大到小的顺序排列。如果该数据的个数为奇数，位于中间的那个数就是中数。如果该数据的个数为偶数，那就取其中间两个数据的平均数。中数是集中趋势的一种表达形式，计算比较简单。当一组数据数量小，或一组数据为等级变量时，一般用中数来作为该数据的集中趋势。

平均数是最常用的一种描述集中趋势的方法。它比众数和中数更为准确，因为它考虑到所有的数据，不易受极端数据影响。其公式为：

$\bar{X} = \sum x / N$

在上述公式中，$\bar{X}$ 代表平均数，$\sum$（读作 sigma）代表几个数字连加，即“……之和”，x 代表每个数据的值，N 代表数据的个数。换句话说，要计算平均数，把所有的数据相加，然后再除以该数据的个数。“在任何情况下，平均数是最好的、也是最实用的和最有用的测量集中趋势的方法”（Hatch & Farhady，1982：55）。然而，平均数不适合称名变量和等级变量。

3. 离散趋势

尽管集中趋势可以很好地描述一组数据的特征，但仅有这些统计量还是不够。我们还需要考虑离散趋势。例如：有两个班平均分数是相同的，是否说明这两个班成绩相同？显然不一定。这些分数的离散趋势，即偏离中心点的程度可能是不同的。因此，为了更为准确地讨论数据，我们不仅要测量其集中趋势，而且要测量其离散趋势。这里我们介绍3种描述离散趋势的方法：全距、标准差和方差。

3.1 全距

全距是说明数据离散程度最简单的统计量。把一组数据按从大到小的顺序排列，用最高分减去最低分，所得到的就是全距。全距大说明数据的分布较为分散，全距小说明数据的分布相对集中。全距的优点是计算方便，而且容易理解。缺点是容易受到极端数据的影响，例如：你班上的学生因为有一个学生考了0分，会使全距急剧加大。因其测量不够精确，也不够稳定，因此一般不用在正式的研究中。

3.2 标准差

在正式研究中，使用得最多的测量离散趋势的方法是标准差。在解释标准差之前，我们先来看一个例子。英语期中考试结束后，教师告诉学生这次考试的平均分是82。很自然，各个学生所关心的是自己与平均分的差距。假如学生甲考了90分，这样，他比平均分高8分；学生乙考了65分，比平均分低17分。作为研究人员，我们所关心的不仅仅是个体偏差，还有所有分数的平均偏差。用“标准”这个词，是因为要看在平均分周围所有值的平均离散趋势。那么，如何计算标准差呢？其公式是：

$$S=\sqrt{\frac{\sum (X-\overline{X})^2}{N-1}}$$

其中X是一个原始数据，$\overline{X}$是平均数，$\sum (X-\overline{X})$表示将每个数据减去平均数，然后平方，再将其连加起来（即求平方和），再用平方和除以（N－1），最后求平方根，便可得到标准差。

具体计算时，可按照以下5个步骤：

1）计算平均数；

2）拿每个数减去平均数，得到每个数据偏离平均数的值；

3）把每个数据偏离平均数的值平方，再将其连加起来，即求平方和；

4）拿平方和除以 N－1；

5）最后把得数开平方。

请看下面的例子：10 个学生听写的分数是：10，9，9，8，7，7，6，6，5，3，平均分是 7 分，下面这个表说明如何得到标准差。

X（各人成绩）	x（与均分的偏差）	x^2
10	3	9
9	2	4
9	2	4
8	1	1
7	0	0
7	0	0
6	－1	1
6	－1	1
5	－2	4
3	－4	16
$\sum X = 70$ N = 10 Mean = 7	$\sum x = 0$	$\sum x^2 = 40$ $\sum x^2/N-1 = 4.44$ SD = 2.11

表 10.6：10 名学生成绩的标准差

标准差是什么？它是一组数据所有值与平均数的平均偏差或离散程度。标准差的值越大，说明离散程度越大，标准差的值越小，说明数据之间的差异越小，或者说离散程度越小。假如你是新来的英语老师，系主任告诉你这儿有 3 个三级班，你可以从中挑选一个。表 10.7 是上学期这 3 个班期终考试的平均分和标准差。

班级	平均分	标准差
一班	83.4	5.2
二班	82.3	5.0
三班	84.1	8.4

表 10.7：3 个班考试的统计量描述

根据上表提供的信息，如何作出选择？如果只看他们的平均分，三班最高，二班最低，但是 3 个班的差异不大。接着，看这 3 个班的标准差，你马上会发现三班的标准差比一班和二班大得多。你想教哪个班？想教成绩参差不齐的班级，还是成绩差不多的班级？标准差越大，班上

的成绩差异越大，标准差越小，班上的成绩差异越小。从这个意思上说，标准差可以提供平均数所不能提供的信息。因此，在我们数据分析时，既要计算平均数，也要计算标准差。

3.3　方差

方差是测量离散趋势的方法之一，用 V 或 S 表示，就是标准差不开方。其公式为：

$$S^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

你可能已经注意到了标准差是方差的平方根。因此，要计算标准差，首先必须计算方差。

第三节　推断性统计

描述性统计学只描述样本数据的频率、集中趋势和离散程度。推断性统计学不仅需要描述数据，而且要预测样本在多大程度上可以代表总体。在我们讨论推断性统计学的细节之前，我们必须解释两个重要概念：正态分布和概率。

1. 正态分布

在从理论上解释什么是正态分布之前，我们先来画两组数据的频率分布。

第一组数据：

80，81，81，82，82，82，83，83，84

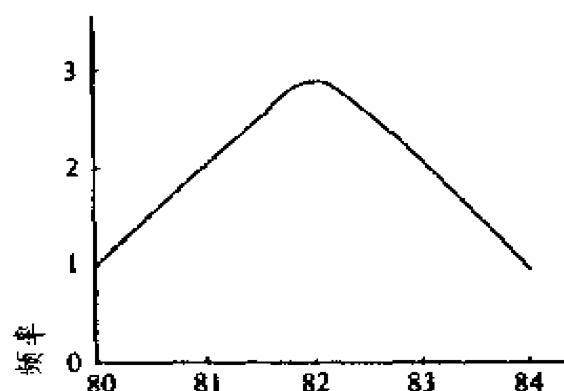


图 10.4：第一组数据的曲线图

第二组数据：

100，99，99，99，99，98，98，98，98，98，98，98，98，98，98，98，98，98，98，98，98，97，97，97，97，96

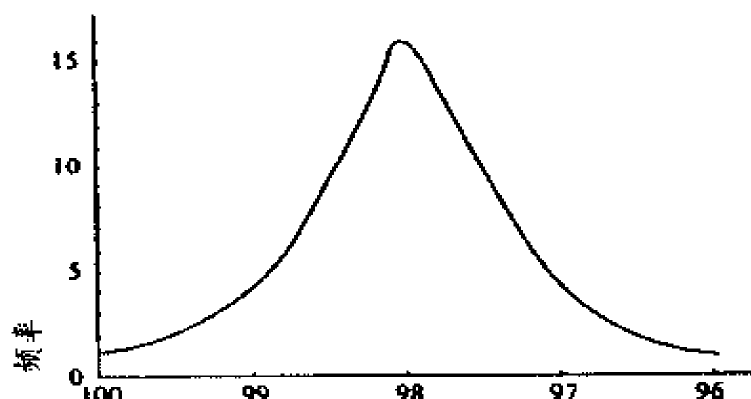


图 10.5：第二组数据的曲线图

通过画这两个曲线图，你可能对正态分布的形状有个大概的了解，整个形状象一个对称的铃铛。你一定还注意到这两个曲线图有点不同：图 10.4 的高峰较低，图 10.5 的高峰较高。然而，这两个图都是正态分布，因为它们都具有下面 4 个明显的特征：

1）它只有一个高峰，即只有一个最高点，看上去“中间高，两边低”；
2）众数、中数和平均数都相同；
3）全距对称，最低值和最高值到平均数的距离相等；
4）正态分布曲线的两个尾端无论怎么延伸，只能是越来越接近 X 轴，但永远不会与 X 轴相交。换句话说，Y 轴的值永远不会为零。

正态分布是描述随机现象的一种理想分布。样本越大，所获得的分布曲线就越靠近正态分布。

2. 概率

什么是概率？它是指事件发生可能性的大小。这里，我们先来看几个例子。每年春节前，各商家为了促销，往往搞有奖销售。设想一等奖为住房一套，25 万张奖券设一个一等奖，那么，每一张奖券获一等奖的概率为 25 万分之一，因此其获奖的可能性几乎为零。再如，考试中的选择题，如果是四选一，那么其正确的概率为 25%，如果是正误判断

题，其正确的概率为 50%。

概率的公式为：

P = n_e（理想事件的次数）/N（可能结果的次数）

理想事件的次数 n_e 永远不会大于可能结果的次数 N。概率的最大值是 1。如果某事发生的概率是 1 时，这就意味着这件事发生的可能性为百分之百。概率的最小值是 0，这就意味着这件事百分之百地不会发生。因此，任何概率的值都不可能是负数或者大于 1，永远是在 0 与 1 之间。概率的值还可以通过乘 100% 转换成百分数。例如，我们可以说正误判断题的正确答案的概率是 .50，我们也可以说正误判断题的正确答案的概率是 50%。

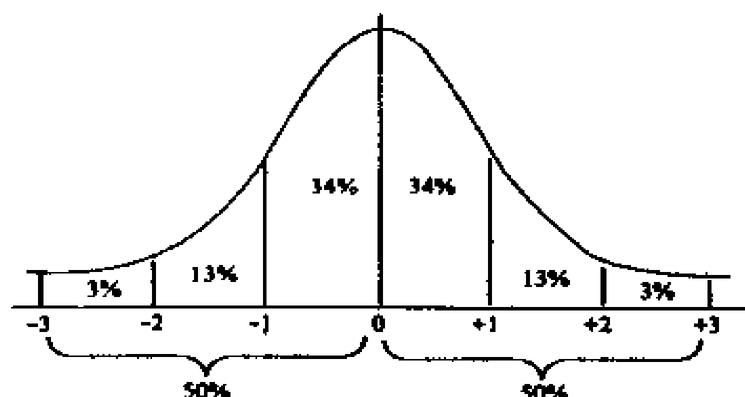


图 10.6：正态分布下标准差与概率的比率

3. 正态分布与概率的实际使用

正态分布和概率可帮助我们确定下面两个问题：

1）某个分数出现的概率是多少？

2）所选样本能代表总体吗？

由于正态分布是对称的，其平均数、中数和众数三者相等。因此一半的数字高于平均数，一半的数字低于平均数。在图 10.6 中，数学家们进一步描述得到某个分数的概率（Nunan，1994：31）。

根据上图，我们可以说得到 ± 1 个标准差以内的分数占全部分数的 68%（34% + 34%）。得到 ± 2 个标准差以内的分数的概率为 95%。得到 ± 3 个标准差以内的分数的概率为 99%。

了解有关概率和正态分布的情况，我们可能会知道所研究的样本是

否代表总体。如果样本可以代表总体，那么，样本的平均数与总体的平均数就没有显著差异。我们什么时候可以说样本的平均数与总体的平均数没有差异呢？实际上，我们在检验零假设。

如果样本在95%的概率区域，那么，样本与总体相同，这样，就不可能排除零假设。

如果样本在5%的概率区域，那么样本与总体不同。换句话说，我们有95%的把握说样本与总体有显著差异。

问题是我们往往得不到总体的平均数和标准差。我们所能做的，就是想方设法得到尽可能多的样本平均数，然后根据样本平均数来估计总体平均数。

4. 检验假设的过程

在研究中，人们总是想发现有关研究结果假设的概率。要想做到这一点，我们在提出假设时，必须把样本的数据与总体的数据进行比较。具体地说，需要先提出零假设和备择假设。接着决定排除零假设的显著水平。在应用语言学领域，显著性水平通常是 .05、.01 和 .001。

设想我们决定了推翻零假设的显著性水平为 .05，如果样本平均数落在95%的区域（如图10.7a所示），零假设就不能被否定。这是因为我们所比较的平均数能够很容易在正态分布95%区域内找到；如果落在 .05 的区域（如图10.7b所示），就要否定零假设，这是因为我们比较的平均数不是正态分布中的典型分数。如果这个分数在左侧，就说明样本平均数不如总体平均数高，如果在右侧，就说明样本平均数比总体高。无论在左侧，还是在右侧，我们都要否定零假设。

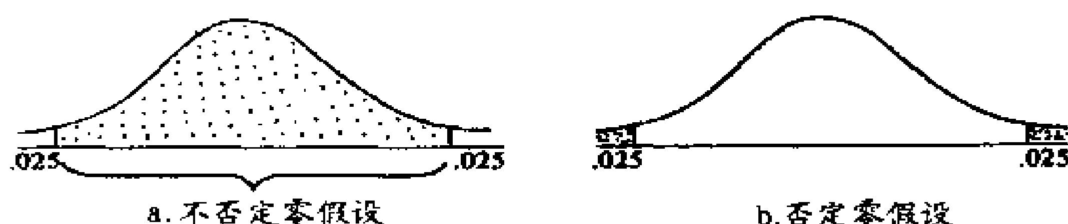


图10.7：否定零假设或不否定零假设
（Hatch & Farhady，1982：87）

同样，如果所选的显著性水平是 .01，当样本平均数落在99%的区

域，就不能否定零假设。如果所选的显著性水平是 .001，当样本平均数落在 99.9% 的区域时，就不能否定零假设。

在图 10.7 中，我们选择了 .05 作为显著性水平。由于正态分布有两个尾段，因此 .05 需要除以 2，结果为 .025。这叫做双尾检验。

小结

描述性统计学和推断性统计学是统计学的两个分支。描述性统计学描述一组数据的全貌，揭示一个事物的总体特征，如：频率、集中趋势和离散程度。推断性统计学要根据样本的统计量来推断总体的情况。参数用来描述总体，统计量用来描述样本。当总体与样本有显著差异时，这一区分非常有必要。参数用希腊字母表示，统计量用罗马字母表示。在推断性统计学中，有两个重要概念：正态分布和概率。有了这两个概念，我们可以检验零假设。如果得到相关结果的概率小于 .05，就可以否定零假设，相应地接受备择假设。这种概率水平又称作为显著性水平，在应用语言学和许多社会科学领域，显著性水平通常是 .05、.01 和 .001。

讨论题

1. 描述性统计学与推断性统计学有什么差别？
2. 下面是 20 个英语专业二年级学生的四级考试的分数：50，60，70，75，55，60，65，70，80，85，55，60，80，80，65，45，70，65，70，75
   1）把这些分数分成 3 组，然后用表和图来表示。
   2）把这些分数分成 4 组，然后用表和图来表示。
3. 下面是 20 个英语专业四年级学生八级考试作文的分数：12，14，10，12，16，18，10，12，16，18，14，12，16，10，12，14，10，18，10，12
   1）计算出这些分数的平均数、中数和众数。
   2）计算出这些分数的全距、方差和标准差。

第十一章　定量数据分析

在前面的讨论中，我们说过，定量数据分析有两种：一种是描述性统计方法，另一种是用推断性统计方法。使用哪种方法主要取决于研究目的。本章首先对 SPSS 统计软件包做扼要的介绍，然后重点讨论如何分析问卷调查和实验研究中的数据。

第一节　SPSS for Windows 概述

SPSS 是一个综合性和灵活性很强，且容易操作的统计分析和管理系统（Norusis，1993）。运行这个软件系统，只要轻松点击一下鼠标，就能完成多种数据分析任务。这节将对 SPSS 软件包作一个概述。

1.SPSS 窗口



图 11.1：SPSS 数据编辑窗口

启动 SPSS 程序后，我们可以看到以下数据编辑窗口，如图 11.1 所示（SPSS 11.5 for Windows）：

数据编辑窗口的界面类似 Word 的操作环境，上面有菜单选项。运用窗口可以创建新的文件，也可以用来打开已有的数据文件，对数据进行各种处理和分析。

2. 数据编辑窗口上的菜单介绍

Windows 环境下运行的 SPSS 操作简便，各菜单选项在运用窗口顶部一目了然，主菜单条上共有 10 个菜单项：

文件（File） 类似于 Microsoft Word 中的菜单，可用于新建文件或打开已经储存的文件，也可以保存编辑后的旧文件或新建文件。

编辑（Edit） 也与 Microsoft Word 相似，主要用于修改或拷贝输出窗口中的文本。

视图（View） 根据需要决定屏幕上显示的内容。

数据（Data） 用于定义变量、插入变量、合并文件、分离文件、选择分析案例等。

转换（Transform） 可以修改数据文件，如可以根据已有变量的值计算出新的变量，也可以将一个变量转换成新的变量，或通过改变一个变量的值对该变量进行重新编码。

统计（Analyze） 可以根据自己研究的需要选择不同的统计程序，如相关系数、比较均值、线性回归等。

图表（Graphs） 主要用来制作各类图表。

功用（Utilities） 可以用来改变字体、显示各变量的相关信息，也可以用来打开 SPSS 命令索引。

窗口（Window） 用来编排、选择和控制各个 SPSS 窗口，如可以将数据编辑窗口转成输出窗口。

帮助（Help） 提供信息，以帮助说明如何使用不同菜单项的窗口。

第二节　问卷数据分析

这一节将以文秋芳（2001）的一项研究* 为实例，详细说明问卷数据分析的方法。

1. Wen 的研究简述

本研究调查英语专业二年级学生学习英语的方法，并分析他们的学习方法与学习结果之间的相关程度。

1997 年 9 月，南京大学 3 个自然班级的 77 名英语专业二年级学生参加了该项研究。他们在上课时间回答了“语言学习者因素调查问卷”，回答率是 100%。77 名受试对象中，有 59 名女生，18 名男生。年龄跨度是 18 至 20 岁，平均年龄为 19.4 岁。1998 年 5 月他们参加了英语专业四级水平测试。

“语言学习者因素调查问卷”包括 3 个部分：第一部分包括两个方面的内容。第一个内容包含受试对象的个人背景情况（如姓名、性别、出生年月、英语及语文高考入学成绩以及作为“学习努力”指标的课外学习英语所花的大概时间等）。第二个内容是调查学生学习英语的原因，答卷人要从 5 分量表列出的 5 个选项中选择一个，即从“1）这个原因对我不重要”到“5）这个原因对我极其重要”之间 5 选 1 作答。第二部分调查学习者的语言学习观念，答卷人仍然是从 5 分量表列出的 5 个答案中选择 1 个，即从“1）我坚决不同意这个看法”到“5）我非常同意这个看法”之间 5 选 1 作答。第三部分调查学习者的学习策略，一部分问卷题涉及管理策略，一部分问卷题涉及语言学习策略。答卷人要在每个问卷题后面的 5 项选择中挑选一个，即从“1）这个陈述完全不或几乎完全不适合我的情况”到“5）这个陈述完全或几乎完全适合我的情况”之间 5 选 1 作答。

* 这里描述的内容是这个项目的一部分。

变量名称	变量内容介绍	题目数	Alpha值
1）英语成绩	英语专业四级测试分数	1	
2）性别	生理性别	1	
3）母语水平	中文高考入学成绩	1	
4）第二语言水平	英语高考入学成绩	1	
5）努力程度	自我估算的课外英语学习时间	1	
6）学习目的	学习英语的理由	3	
7）努力观念	对语言学习中需要下工夫的看法	5	.51
8）管理观念	对规划、设定目标、评价学习重要性的看法	5	.75
9）传统观念	对重复、记忆、精读课文重要性的看法	4	.59
10）非传统观念	观念对广泛接触目标语言、在交际中使用目标语言重要性的看法	4	.64
11）使用母语观念	对在听说读写中使用母语是否必要的看法	4	.69
12）管理策略	在规划、目标设定、评价中的行为	8	.79
13）传统策略	传统学习活动中的行为	11	.73
14）非传统策略	非传统学习活动中的行为	12	.75
15）使用母语策略	在第二语言学习中使用母语的行为	5	.69

表11.1：15个变量的描述

在表11.1中共有15个变量，其中英语成绩是因变量，其余皆为自变量。前5个变量只有一个指标，没有必要进行统计核对。剩下的10个变量，每个变量有多个指标，需要设计多个问题才能获得，而且它们之间必须进行内在一致性检验。表中的Alpha值就是表示多个指标的内在一致性。（注：Alpha值的含义将在“范畴的确定”一节中讨论。）

2. 建立数据文件

问卷收集后，下一步任务是建立数据文件。具体地说，要给数据编码，对问卷进行逐一编号，输入数据，整理数据。

2.1 数据的编码

数据编码不是研究中的重要阶段，但是不可缺少。例如，一个受试者是男性还是女性不会自动转换为能够被统计软件识别的计算形式。研究者必须选择相应的数字代码代表受试者的性别，比如用 1 代表男性，2 代表女性。这里的数字代码只是统计软件可以记录、储存、处理和检索的另一种数据。使用什么样的数字编码取决于测量变量的方法。对于称名数据，使用的数字代码可以是任意的，但对于等级和等距数据，编码的数字应有数学价值，不能随意编码。

编码过程中，最好准备一本代码手册。一本好的代码手册，对日后的数据处理大有裨益，它可以告诉你数据中包含的内容，如变量数目、变量名称、数据文件中不同数字代码的含义等。没有像样的代码手册，其他研究者看不懂你的数据，甚至几个月后连你自己也看不懂，人的记忆力毕竟有限，记错或遗忘的情况会经常发生。

代码手册应该如何编写？表 11.2* 是文秋芳（2001）在研究中使用的代码手册中的一部分：

变量的全称	编码后的变量名	数值的含义
Sex（性别）	Sex	1＝男性；2＝女性
Learning Purpose（学习目的）	LP	1＝不重要；2＝比较重要；3＝重要；4＝非常重要；5＝极其重要
Management Belief（管理观念）	MagBEL	1＝坚决不同意；2＝不同意；3＝中立；4＝同意；5＝非常同意

（见下页）

* 在论文里这种表格通常作为附录放在最后。

（接上页）

变量的全称	编码后的变量名	数值的含义
Using – Mother – Tongue Belief（使用母语的观念）	UMTBEL	1＝坚决不同意；2＝不同意；3＝中立；4＝同意；5＝非常同意
Management Strategy（管理策略）	MagStr	1＝完全不或几乎完全不适合我的情况；2＝通常不适合我的情况；3＝有时适合我的情况；4＝通常适合我的情况；5＝完全或几乎完全适合我的情况

表 11.2：代码手册范例

代码手册应包括下面 3 个方面的内容：

1）对各个变量的完整描述；

2）编码后的变量名称；

3）每个变量可能含有的数值意义。

2.2　问卷的编号

在输入数据之前，应该将收集到的所有问卷进行编号。也就是说，每一份问卷都应该有自己的代码。为了方便，最好使用连续的数字，如 1，2，3，4，5 等。输入数据的时候，这些数字作为每个答卷人的身份号编入第一栏。到了检查数据的时候，如果发现数据有误，这些编号能帮助你找到原始的问卷。

2.3　数据的输入

有些研究者使用标准化答题卡记录研究对象的答案，这样的数据可以用机器读入。但中国的研究生使用人工输入数据的情况更为普遍。下面我们将介绍如何使用 SPSS 中的“数据编辑”窗口来人工输入数据。

启动 SPSS 程序后，屏幕上会出现“数据编辑”窗口，窗口中有一张电子表格或称工作表，表格中，行是案例，列是问卷题项。每个单元格显示出一个案例中某个变量的一个值。

数据输入时首先应该定义变量。点击 11.1 窗口底部的“Variable View”，就得到图 11.2。



图 11.2：“Variable View”

下面是定义每个变量，需要逐项完成：

确定每个变量的名称（Name） 变量的名称必须以英文字母开头，剩下的可以是字母，也可以是数字。所有变量的名称不能超过 8 个字符，变量的名称不能重复。最后需要注意的是，变量的名称应便于辨认。比如：ID 代表身份号，P01 指 Learning Purpose 部分的第 1 个问卷题，B01 指 Belief 部分的第 1 个问卷题，S10 是指 Strategy 部分的第 10 个问卷题。

确定变量的类型（Type） SPSS 只默认数字数据，所有新的变量都被看作是数字。实际上我们的数据都是数字的，因此不需要运用任何程序。

设定变量的位数（Width） 这里的默认值为 8 位数，一般情况下，可以不作改动，除非你研究中所需的位数超过 8 位。

设定变量保留小数的位数（Decimals） 这里的默认值为 2 位数。你可以根据需要增加或减少保留小数的位数。

描写完整的变量名（Labels） 给出每个变量名完整的名称，以备日后查考。

标注每个变量值的意义（Values） 给出每个变量值的意义，例如 female

= 1，male = 2。如果你的代码手册可以帮助你查到各个变量的相关信息，这一步你可以不做。

确定缺值（Missing） 有时我们会碰到某个案例中应答者没有回答问卷中某个问题，或者给出的答案是模糊的，这时我们只能把它看作缺值（missing values）。尽管有些变量也许不会出现缺值，但我们必须假定所有的变量都有缺值，而且指定用固定的数字代表缺值。一般来说，我们可以用“0”来代表缺值，但是“0”有时也是有意义的，比如测试的分数值中“0”是有意义的，这时我们可以用“999”来代替。

设定列宽（Columns） 每列的宽度最少是一个字母，默认值为 8 个字母。如果你设定的变量长度超过 8 位数，这里列宽也要做相应的调整。

确定变量值对齐的方式（Align） 变量值对齐的方式有 3 种：左对齐，右对齐，居中。默认值为右对齐。

描写变量测量的方式（Measure） 变量测量的方式有 3 种：称名测量、等级测量、等距测量。默认值为等距测量。

对所有的变量进行定义之后，下一步就是按照案例输入数据。在输入数据的过程中，最好请一个人在边上帮你朗读这些数字。必须注意的是，输入数据时，中途要注意休息，以免过分劳累，输入准确率不高。有些人急于完成任务，中途不愿意休息，其实疲劳时输入数据，最容易出错，而纠正错误费时费力，其结果是欲速则不达。此外，在开始分析数据之前应对数据文件做一个备份，原始数据文件应与工作数据文件分开储存。

2.4 数据的清理

清理数据主要有两个任务：一是检查数据输入中有无差错，二是处理缺漏的答案。出现差错的数据必须得到纠正，原因很简单，不正确的数据值会直接影响统计分析的结果，甚至会破坏整个数据的分析。有些错误是很容易发现的，比如，忘记了处理缺少的数据值，使用了无效的代码，或把 25 岁键入为 525 岁，这些错误从频数表上很容易看出来。使用莱克特 5 分量表，只可能有 1 到 5 这 5 种答案，如果数据文件中出现了 14，45，12 等数字，那肯定是输入错误。有时候，某个数字倒不是没有可能，只是不太符合常理。例如，在回答“你每周课外学习英语的时间是多少小时?”的时候，如果一个学生在问卷上写下了 100 小时，这个数字不是完全没有可能，但让人产生怀疑，应该检查输入是否有误。还有一类是输入上的错误，很难找出来，如将年龄值 26 当成了 62，3 当成了 5。

问卷调查中出现缺值的情况有两种。一种是被调查者没有回答某道题，因此这道题没有答案。第二种情况是，虽然被调查者选择了答案，但答案不符合要求，例如题目要求只能选择一个答案，他却选择了两个。这时研究者不能判定输入那一个答案为好，只好作为没有答案处理。如果缺漏的数据低于20%，可以使用下面两种办法补救：

使用平均数替代。用平均数替代缺值，数据分析可以不受影响。例如，一份问卷中的2，12，8，10和23题归为一类范畴，可是有两个受试者没有回答第2题，这时第二道题就出现了缺值，这样一来，在形成范畴的过程中，SPSS程序就会自动将这两位问卷对象从样本中剔除出去。但是我们如果使用均值替代这个项目的缺值，SPSS程序会将这两位问卷对象和其他问卷对象一样对待。

成对删除。在成对删除中，每次计算中相关的案例只是那些有完整答案的案例。换言之，那些对问卷中的某个项目没有作出回答的案例在分析该项目的时候就被删除。因此，在每次计算过程中涉及的案例会因完整答卷的数目不同而不同。但这个方法只适合于比较大的样本。

3. 类别的确定

在“问卷设计程序”一节里，我们提到了有两类问卷：以单个项目为基础的问卷和以类别为基础的问卷。在第二种问卷里，需要使用统计程序确定逻辑上可以接受的类别。具体地说，类别内的所有项目必须是同方向，并且要接受内部一致性的检验。

3.1　对相反方向的回答进行重新编码

问卷中的项目呈相反方向，是指这些项目在内容上相互矛盾。我们看下面两个句子：

1）在第二语言学习中，准确性比流利性更重要。

2）在第二语言学习中，流利性比准确性更重要。

确定类别，首先要保证一个类别内的所有项目同方向，即意义上不互相矛盾。如果方向相反，就要改变某些项目的数值，使之方向一致。在文的问卷里，有一个“努力观念”的类别，它表示学生对在第二语言学习中“努力”重要性的认识。该范畴包括以下五个陈述项，要求学生对每个项下的5分量表作出选择（1＝坚决不同意；2＝不同意；3＝中立；4＝同意；5＝非常同意）。

(1) 学外语需要下苦工夫。

(2) 有语言天赋的人不需要下苦功也能学好外语。

(3) 一个人英语学得好主要是他有语言天赋。

(4) 如果不用功，中国没有人能学好外语。

(5) 学习外语，后天的努力比语言天赋更重要。

我们检查一下前3项：

(1) 学外语需要下苦工夫。

(2) 有语言天赋的人不需要下苦功也能学好外语。

(3) 一个人英语学得好主要是他有语言天赋。

1	2	3	4	5
坚决不同意	不同意	不知道	同意	完全同意

从逻辑上讲，如果你坚决不同意第一项，那你就该非常赞同第二、三项。也就是说，第一项选1就等于第二、三项选5，反之亦然。从这个意义上讲，第一项和第二、三项不是同一个方向。

针对这种情况，我们该怎么办呢？既然这5个问卷题构成的类别叫做“努力观念”，那么对第2、3项的回答就得重新编码。1需变成5，2变成4，5变成1，4变为2，这样一来所有的项目就在同一个方向上。换句话说，它们都强调“努力”的重要性，否则，这些陈述项就会相互冲突，因为第1、4、5项强调“努力”在学习语言中的作用，而第2、3项强调语言天赋的重要。

那么如何给变量重新编码呢？很简单，只要双击 Transform（转换）菜单项，就可以找到 Recode（重新编码）。这里有两个选项：Recode into the same variable（重新编码为相同变量）和 Recode into a different variable（重新编码为不同变量）。安全的做法是选择第二个选项，因为选择第一选项一旦出错就会丢失数据。

3.2 通过个别一总体相关性检查内在一致性

下一步是通过个别一总体相关性来检查类别内部各项目是否具有内在一致性。什么是个别一总体相关呢？我们还是先看一个例子。“管理观念”是一个假设的类别，包含5个陈述项，我们将这些项目在77名英语专业二年级学生中做了调查。

1) 合理安排学习时间是成功的重要因素。

2) 选择合适的学习策略是学好外语的重要因素。
3) 设定学习目标和评价学习进展是学好外语的重要因素。
4) 经常回顾学习英语的进展以找出自己的弱点是学好英语的重要因素。
5) 思考自己的学习策略是否有效是学好外语的重要因素。

文秋芳的研究项目中的数据（如表 11.3 所示）:

	项目 1	项目 2	项目 3	项目 4	项目 5
学生 1	X	X	X	X	X
学生 2	X	X	X	X	X
学生 3	X	X	X	X	X
学生 4	X	X	X	X	X
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
学生 77	X	X	X	X	X

表 11.3：数据文件描述

在表 11.3 中，X 是指每个学生在每个项目上的得分。5 个项目中，分值都是从 1 到 5。最低的可能得 5 分，最高的可能得 25 分。当然，在调查中，每个答卷者总分可能介于两者之间。有了这 5 个问卷题，答卷者的总分能帮助你了解每个学生在你所要测量的这个类别所处的位置。接着，你就可以运算出各项个别—总体的相关性。这里的总体不是某个受试者的个人得分，而是 77 个学生的总分，同样，每个项目的得分也不是个人的得分，而是 77 个学生的得分。

要运行个别—总体的相关性，打开 Statistics（统计）菜单项，找到 Scale（量表），点击 Scale 后有两个选项供你选择：(1) Reliability Analysis（信度分析），(2) Multidimentional Scaling（多维测量）。计算个别—总体相关性，需要点击第一个选项。表 11.4 是信度分析结果。表中的 Alpha 值高达 .7542，所以假设的“管理观念”范畴得到了确认。换言之，该范畴可以用来做进一步统计分析。

Item – total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item – Total Correlation	Alpha if item Deleted
Item 1	17.3247	3.5906	.4758	.7258
Item 2	17.0000	3.4474	.6147	.6833
Item 3	17.4156	3.3513	.4343	.7472
Item 4	17.2468	3.1883	.6588	.6607
Item 5	17.4545	3.3301	.4658	.7332

Reliability Coefficients

N of Cases = 77.0　　N of Items = 5

Alpha = .7542

表 11.4：管理观念范畴的项分值—总分值相关

所有假设的类别都需要经过上述统计确认。有时候我们没有那么走运。尽管在逻辑上类别内部的各项应该互相关联，但这样符合逻辑的类别不一定能得到实证性数据的支持。有时只需要删除一些项目或对某些项目进行重新组合或排列。糟糕的时候，整个类别必须舍弃。如何对类别内部的项目作修改呢？我们可以观察最后一栏“Alpha if deleted”（删除项目后的 Alpha 值），就可以知道哪一个项目舍弃对类别内在一致性影响最大。在“管理观念类别”这个例子中，我们不需要做任何修改，因为无论删除哪一项，都会降低内在一致性。我们再看一个例子（如表 11.5 所示）。在这个例子中，有四个问卷题构成了“使用母语观念”类别。这四项涵盖了学生对在第二语言听、说、读、写中使用母语的看法：

1）要很好地理解课文，最好的办法是翻译课文。

2）记住所听内容的最好的办法是使用汉语来记住所听内容。

3）要想写出好的英语作文，最好的办法是用母语来组织作文的内容。

4）说英语的时候，最好的办法是先用汉语思考一下你想说的内容。

按照表 11.5 最后一栏“Alpha if deleted”，如果项目 4 删除，alpha 值就会增加到 .6661。与之相反，如果删除项目 3，alpha 值就会下降到 .4988。决定每个项目的取舍不能只看统计数据。实际上，alpha 值只是一个方面的因素。就“使用母语观念类别”而言，删除项目 4 也许会提高 alpha 值，

但这个项目描述的是在第二语言口语中使用母语的重要性，这是第二语言学习的特定方面，因此不宜删除。当然，假如保留项目 4 会导致 alpha 值过低，以致影响了整个类别的信度，那就应毫不犹豫地删除这个项目。

Item - total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item - Total Correlation	Alpha if Item Deleted
Item 1	12.7013	1.7912	.3891	.5902
Item 2	12.7143	1.6015	.4770	.5278
Item 3	12.5325	1.6733	.5401	.4988
Item 4	13.2078	1.5615	.3195	.6661

Reliability Coefficients

N of Cases = 77.0　　N of Items = 4　　Alpha = .6386

表 11.5：使用母语观念范畴的项分值—总分值相关

除了类别信度（Scale Reliability）这个统计程序之外，我们还可以使用更精确的统计程序建立或确立类别，这就是因子分析。假设你的问卷有 30 条题目，准备将它们归为 5 个类别。因子分析的前提是这些类别必须相互之间不相关，或相关的程度非常低。如果你假设的 5 个类别能够符合这个前提条件，因子分析的结果就是 5 个不同的因子，且每个类别内部的各项目都有很高的因子负载。然而，现实比理论要复杂得多。建立类别，首先要设计大量的问卷题目，请不同的群体对各个题目作出应答，然后根据不同受试者的回答重复运行因子分析。每次因子分析之后，只有那些因子负载高的项目保留下来。换句话说，必须通过因子分析重复检验这些类别，在这个过程中，一些因子负载不足的项目就被剔除出去。

虽然使用因子分析的方法建立类别比检查项目内在一致性更加精确，但有些描写第二语言学习过程的类别不能确立为不同因子，因为这些类别从理论的角度来看应该彼此独立，互不关联，但在现实中彼此交叉，互相联系，比如在 Wen 的研究项目中，“观念”和“策略”理论上是两个不同的类别，但在现实生活中，两者密不可分。类似这样的情况，我们就只能用类别信度的方法来确立类别。

4. 选择合适的统计方法

可供选择的统计方法很多，但是哪一种最适合分析你的研究问题？在选择统计方法的过程中，需要作出一系列的决策。在下面一节里，我们将介绍 SPSS 中 Statistics 菜单项中的最常见的 4 种统计程序：Summarize（概括），Compare means（比较均值），Correlate 和 Regression（回归）。有一点必须强调的是，除了 Summarize 外，所有统计程序的运行都必须以正态分布的假设为前提。

4.1 概括：频数统计和描述性统计

打开 Statistics 菜单项，会出现很多选项，最上方是 Summarize 选项。由于篇幅所限，这里我们只对前两项作简要的介绍：Frequencies（频数统计）和 Descriptives（描述性统计）。在撰写硕士论文或博士论文过程中，这两个选项是远远不够的，通常需要更复杂的统计分析，但它们是统计分析的第一步。使用 Frequencies 和 Descriptives 可以帮助你对数据有一个初步的总体了解。

例如，要想了解某项研究中有多少学生是女生，有多少是男生，各占比例是多少，运行 Frequencies，很容易得到答案。下面的统计结果清楚地告诉我们 18 名男生和 59 名女生以及她们占学生总数的百分比。

SEX（性别）

Value label	Value	Frequency	Percent	Valid Percent（有效百分比）	Cum Percent（累计百分比）
	1.00	18	23.4	23.4	23.4
	2.00	59	76.6	76.6	100.0
	总计	77	100.0	100.0	

Valid cases（有效人数）77 Missing cases（缺值人数） 0

如果问卷中只有有限的几个问题，那只需汇报频数和/或百分比。

描述性统计能告诉我们平均分和每个范畴的标准差。表 11.6 中的结果就是描述性统计的结果，里面包括 Wen 的研究中“观念”范畴各变量的平均分和标准差。从表中的结果来看，英语专业二年级学生对使用母语的观念持否定态度，而对其他 4 个观念持肯定的态度。但是与传统观念相比，他们更坚持非传统的观念，传统观念的平均分值（3.96）比非传统观念的平均分值（4.27）要低。

此外，表中 UMTBEL 的标准差最大，而 NTRABEL 的标准差最小，这

说明学生对使用母语的看法要比对非传统观念的看法更呈多样化趋势。

Variable Label (变量名称)	Mean (均值)	Std Dev (标准差)	Valid N (有效人数)
UMTBEL (使用母语观念)	2.13	.56	77
EFBEL (努力观念)	3.87	.48	77
TRABEL (传统观念)	3.96	.43	77
NTRABEL (非传统观念)	4.27	.40	77
MAGBEL (管理观念)	4.32	.45	77

表 11.6：观念变量描述性统计

（注：该表是 Descriptives 生成的结果，汉语是笔者加的。）

4.2 相关分析

相关分析是用来检验两个变量相互关系的显著程度。相关分析的结果能让我们知道 3 个方面的内容：

1）两个变量之间的关系如何？换言之，我们在多大程度上能从自变量的分值预测样本中因变量的分值？

2）这种关系是偶然的，还是存在于我们需要推测的整体样本之中？换句话说，这种关系是否具有显著性？

3）两个变量的变化方向如何？呈正相关还是负相关？

所谓变化的方向，指两个变量之间的协同关系是一致还是相反。如果两个变量的变化方向一致，即一个变量值变大时，另一个变量值也随之变大，反之亦然，那么两变量之间的关系呈正相关；如果两个变量的变化相反，即一个变量值变大时，另一个变量值随之变小，反之亦然，那么两变量之间的关系呈负相关。例如，学生英语入学考试的成绩和他们的英语专业四级考试成绩之间的关系呈正相关变化，因此英语入学考试成绩高的学生在英语专业四级考试中往往也会取得好的成绩；相反，入学考试成绩低的学生在四级考试中的成绩也往往会偏低。再如，在英语学习中使用母语与英语专业四级考试的成绩的协同关系是负相关，那么学生在学习第二语言中使用母语越多，在四级考试中的成绩就可能越低，而在第二语言学习中使用母语越少，四级考试的成绩就可能会越高。

总之，正相关是指两个变量朝同一方向变化（如图 11.3a 所示），而负相关是指两个变量朝相反的方向变化（如图 11.3b 所示）。

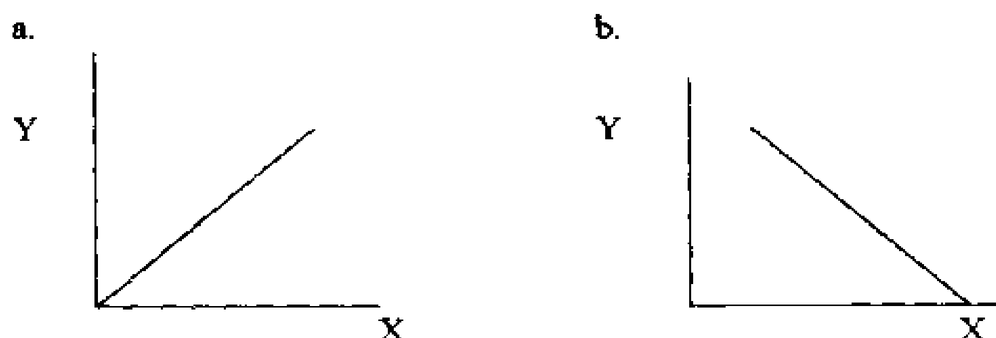


图 11.3（a）：两个变量方向相同的的相关情况
图 11.3（b）：两个变量不同方向的相关情况

两个变量之间的相关程度用相关系数表示，相关系数是两个变量之间相关程度的一个指标，取值介于 0 和 ± 1.00 之间。如果相关系数是 0，那么两个变量之间呈零相关关系，即两个变量值的变化方向没有规律，且变量值变大变小的机会趋于相等。如果两个变量之间完全是正相关关系，那么相关系数就应该是 + 1.00，反之是 − 1.00。不过大多数情况下，相关系数都落在 0 和 ± 1.00 之间。例如，第二语言学习的时间总小时数往往与第二语言学习的成绩呈正相关关系，但有些学生学习非常刻苦，却没有取得好的成绩。因此，大多数情况下，两个变量之间不可能完全相关。

一般来说，如果相关系数达到了 .05，.01 或 .001 显著性水平时，我们就可以说这个相关系数具有统计意义。如果达到了 .05 的水平，统计结果的概率达到了 95%。换句话说，统计的结果中仍旧有 5% 的机会可能出错。同样，显著水平如果达到了 .01，那么统计结果出现的概率就是 99%，这意味着只有 1% 的机会可能出错。如果出错的概率超过了 .05，那所获取的相关系数就被视为没有显著意义。

就测量的等级而言，什么样的变量可以用来检验相关性？一般来说，检验相关性主要是在两个等距变量之间，或一个等距变量和一个等级变量之间，或一个等距变量和一个称名变量之间进行（Bernard，1994）。

运行相关分析的程序，只要打开 Statistics 菜单项，点击 Correlate（相关），然后选择 Bivariate（双元方差），屏幕上就会显示出 Bivariate Correlation 对话框。

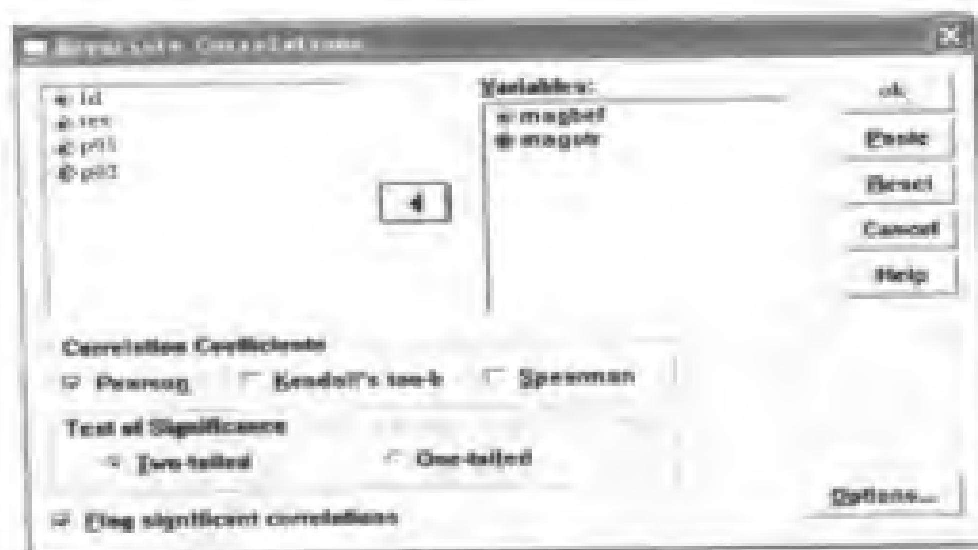


图 11.4：Bivariate Correlation 对话框

具体操作办法是选中左边方框里的两个可能相关的变量，点击中间的箭头，两个变量就会出现在右边标有"Variables"的方框里。表 11.7 显示的结果是 Wen（1993）的研究中学生的观念与使用相应的策略之间的相关系数。在相关分析中，这 4 对变量都被当成等距变量。统计结果表明，4 对变量都呈正相关。但 NTRABEL（非传统观念）和 NTRASTRA（非传统策略）之间的相关系数没有显著意义，因为错误的概率（.062）大于 .05。

变量名称	案例数	相关系数（r）	P
MAGBEL/MAGSTR（管理观念/管理策略）	77	.3763	.001
TRABEL/TRASTR（传统观念/传统策略）	77	.2498	.028
NTRABEL/NTRASTR（非传统观念/非传统策略）	76	.2148	.062
UMTBEL/UMTSTR（使用母语观念/使用母语策略）	77	.5068	.000

表 11.7：观念变量和策略变量之间的相关情况

解释相关分析的结果，须注意两个原则：

1）并非所有概率在 95% 以上的统计结果都具有相同的重要性；2）

如果两个变量之间的相关程度很高，且具有显著意义，这不等于两个变量之间存在因果关系。

4.3 比较平均数：t 检验和单向方差分析

比较平均数的目的是检验差异的显著性。在研究中，我们常常需要比较两个或更多样本的平均数的差异，以此来推断各自代表的总体平均数之间是否存在显著性差异。SPSS 中的“Compare Means”就可以用来比较平均数差异的显著性。其中用得最多的是独立样本 t 检验（independent-samples t-test），配对样本 t 检验（paired-samples t-test）和单向方差分析（one-way ANOVA）。这一节首先说明何时使用上述统计分析方法，如何解释这些统计方法得出的结果，然后介绍如何使用 SPSS 操作这些统计分析方法。

首先看 t 检验的方法。例如，在 Wen（1993）的一项研究中，研究者要调查：1）传统策略的平均数和非传统策略的平均数是否有显著的区别；2）男生和女生在使用非传统策略方面是否有显著差异。要回答这两个问题，就需要使用 t 检验的方法。

t 检验有两种：独立样本 t 检验和配对样本 t 检验。独立样本 t 检验，顾名思义，是检验两组不同群体之间的差异，如男生女生之间的差异、高分组与低分组之间存在的差异，而配对样本 t 检验只检验涉及同一样本中的两个不同变量，如传统策略与非传统策略，内在动机和外在动机等之间的差异。

如果要检验传统策略的平均数和非传统策略的平均数之间有无显著性差异，应该使用配对样本 t 检验的方法。表 11.8 是检验结果。

变量名称	平均值	标准差	T－值	P
Trastr（传统策略）	3.4749	.398	4.21	.000
Ntrastr（非传统策略）	3.2982	.421		

表 11.8：使用传统策略和非传统策略的 t 检验

从上表的结果来看，我们的学生使用传统策略多于非传统策略，且两种策略的差异显著性达到了 .000 水平。

如果要调查男女学生在使用非传统策略上是否具有显著差异，就要使用独立样本 t 检验的方法。从表 11.9 可以看出，女生组的平均数比男生组稍高，但男生和女生之间在使用非传统策略上并没有显著差异。

变量名称	案例数	平均值	标准差	T-值	P
男生	18	3.1713	.458	-1.48	.144
女生	58	3.3376	.405		

表 11.9：男生和女生使用非传统策略的 *t* 检验

独立样本 *t* 检验牵涉到两个独立样本平均数的比较，是对两个样本参加一次测试的两个平均值作比较。配对样本 *t* 检验有别于独立样本 *t* 检验，它是对一个样本参加两次测试的平均值作比较。换句话说，配对样本 *t* 检验操作的对象是相同的一组受试对象，不同的两个变量，而独立样本 *t* 检验的对象是不同两组受试对象所涉及的同一个变量。如果要比较两个以上的样本平均值，那就要选用单向方差分析。假设我们想了解英语专业四级考试成绩不同的学生在使用非传统策略方面有无显著差异，我们该如何入手？首先根据四级考试成绩将学生分成 3 组：高分组，中分组和低分组，然后比较他们使用非传统策略的平均分值。

问题是这些平均数有显著的差异吗？换句话说，就算它们之间有数值上的差异，这 3 组受试对象是否来自同一群体呢？表 11.10 所报告的是单向方差分析结果。判断差异是否具有显著意义，关键看表中 F 比（F Ratio）和 F 概率（F Probability）。因为表中的 F 比的概率是（.0718），大于 .05，因此这 3 组之间的差异没有显著意义。

Variable Non-Traditional Strategy By Variable Band 4 Test Scores Analysis of Variance					
Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	2	1.2501	.6251	2.7372	.0718
Within Groups	69	15.7564	.2284		
Total	71	17.0065			

表 11.10：由 ANOVA 检验的非传统策略使用方面的差异

我们再看一个每组之间有显著差异的例子。在 Xu（2001）的研究中，研究者用方差分析将 4 个年级的学生（即一、二、三、四年级）作文的分数作了比较。

GN	Mean	SD	GP1	GP2	GP3	F	P
GP1（50）	3.17	.56			*	28.39	.000
GP2（50）	3.32	.44			*		
GP3（50）	3.65	.53	*	*			
GP4（50）	3.94	.25	*	*	*		

表 11.11：由 ANOVA 检验的四个组作文成绩差异
（注：* 表示所涉及的两个组有显著差异）

表 11.11 显示，F 值是 28.39，它出现的概率为 .000。这就表明，4 个年级学生写作分数的总体差异具有统计上的显著意义。此外，这个结果还表明，一年级和二年级在分数上没有显著的差异，但这两组学生与三年级和四年级组都有显著的差异，三年级和四年级之间也有统计上的显著差异。

如何利用 SPSS 比较平均数呢？首先打开 Statistics 菜单项，Compare Means 位于菜单项的第三项。这时用鼠标点击，就能看到右边有一组选项，包括 Independent-samples *t*-test，Paired-samples *t*-test 和 One-Way ANOVA。双击 Independent-samples *t*-test，就能获得如图 11.5 所示的对话框。在 Test Variables（检验变量）的下面，从左边的清单上选择你想比较的变量。在 Grouping Variable（分组变量）下面选中代表不同测试组的变量。在表 11.10 的例子中，Ntrstra（非传统策略）应该从左边的清单中移到 Test Variables 的下面，Sex 则移到 Grouping Variable 下面的方框里。

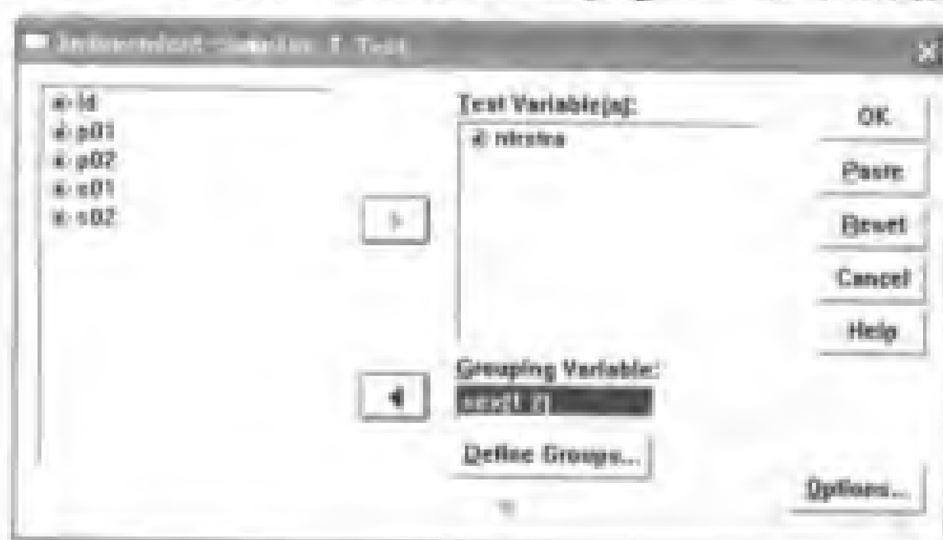


图 11.5：Independent-Samples T-Test 对话框

下一步，双击 Define Groups，就会显示出 Define Groups 的对话框（见图 11.5），在 Group 1 和 Group 2 里分别键入 1 和 2。配对样本 t 检验的具体操作办法是，先点击 Statistics 菜单项里的 Compare Means，右边有一组选项，双击 Independent-Samples T Test，会看到如图 11.7 所示的对话框。在对话框的右下角，Current Selections（目前的选择）下面有两行命名的变量，轮流选择这两个变量进行比较。在表 11.7 的那个例子中，第一个变量是非传统策略（Ntrastr），第二个变量是传统策略（Trastr）。选中这两个变量，将这对变量移到 Paired Variables（配对变量）下面。点击 OK，就可以获得检验的结果。

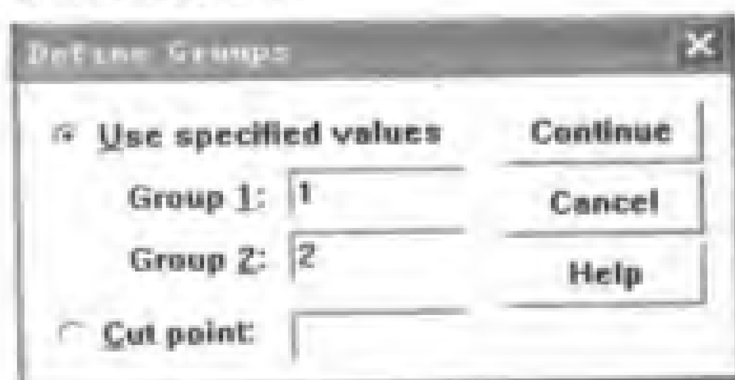


图 11.6：Define Groups 对话框



图 11.7：Paired Samples T-Test 对话框

在进行单向方差分析之前，先点击 Statistics 菜单项里的 Compare Means，右边有一组选项，One-Way ANOVA 在选项清单的最下面。双击 One-Way ANOVA，屏幕上就会出现下面的对话框。

单向方差分析的操作方法分 3 个步骤。第一步，将左边需要比较的变量移到右边 Dependent List 下面。第二步，将代表不同样本组的变量从左移到右边 Factor 的下面。在表 11.8 的例子里，需要比较的变量是分数（Scores），因子是 Grades（年级）。最后点击 Post-Hoc，出现的对话框是 One-Way ANOVA Post-Hoc Multiple Comparisons（单向方差分析事后多样比较）（见图 11.9），然后选中对话框中第一竖行中的 Scheffe。这一步必不可少，否则无法获得多样比较的详细信息。

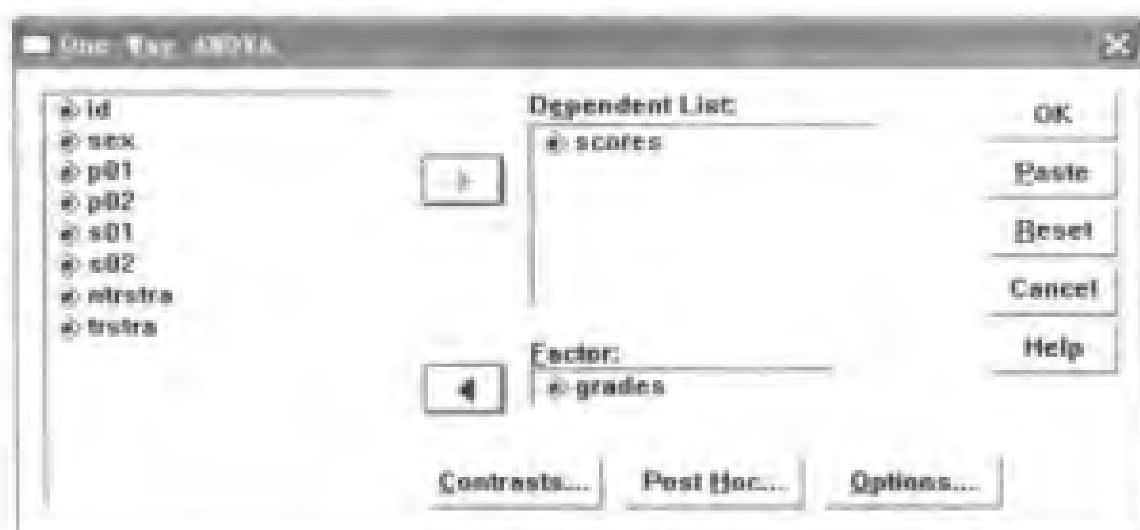


图 11.8：One-Way ANOVA 对话框



图 11.9：One-Way ANOVA：Post Hoc
Multiple Comparisons 对话框（SPSS Inc.：278）

4.4 多元回归分析

回归和相关是两个关系十分密切的概念。相关表示两个变量之间双向的相互关系，如果两个变量之间存在较高的相关，我们就可以通过回归分析，从一个变量的变化去预测或推断另一个变量的相应变化。

所谓多元回归分析，就是测量多个变量之间的多重相互关系以及多个自变量中每个自变量对因变量的预测力。在现实生活中，我们经常作这样的预测，只是没有意识到这是多元回归分析。例如，我们可以根据一个人的性别、年龄和身高去预测他的体重。多元回归分析能够告诉我们性别、年龄、身高各自对体重的预测能力。

在应用语言学领域里，我们常常碰到的问题是，诸如母语水平、入学时的第二语言水平、学习动机、学习努力情况、观念、策略等一系列自变量在多大程度上能够预测学生第二语言的学习成绩？类似这样的问题多元回归分析可以给出答案。表 11.12 是 Wen（1993）在研究学习者可控因素和第二语言学习成绩之间关系时，使用多元回归分析的结果。

Dependent Variable	Multiple R	R^2	F	
Scores on Band 4	.62	.39	2.57 * *	
Independent Variable			Beta	Sig T
1. L1 proficiency			.03	.78
2. Previous L2 Prof.			.43	.001
3. Deep motivation			.23	.06
4. Surface motivation			.11	.41
5. Effort			- .15	.30
6. Management belief			- .14	.32
7. Form-focused Belief			- .13	.36
8. Meaning-focused Belief			.16	.23
9. Mother-tongue Belief			- .01	.95
10. Management Strategy			.02	.88
11. Form-focused Strategy			- .07	.59
12. Meaning-focused Strategy			.24	.06
13. Mother-tongue Strategy			- .32	.03

表 11.12：The Predicting Power of Learner Factors on L2 Learning

（注：* * 意为"< .001"）



图 11.10：Linear Regression 对话框

从表 11.12 来看，13 个自变量能共同预测四级考试（自变量）分数 39%的变异，显著水平达到了 .001。根据 Beta 的绝对值，我们发现最能预测因变量的自变量是“入学时的第二语言水平”，它的 Beta 值（.43）最高，显著水平为 .001。第二个最有预测能力的是“母语策略”，它的绝对 Beta 值是 .32，显著水平为 .03。“深层动机”和“注重意义的策略”具有相同的预测能力，它们的 Beta 值几乎相同（.23 和 .24），显著水平都在 .06。相比而言，它们没有“母语策略”的预测能力强，因为它们的 Beta 值没有后者高，错误概率大于 .05。其他自变量的绝对 Beta 值都很低，也没有达到显著性水平，因此没有预测能力。从表 11.12，我们可以看到，有些变量的 Beta 值是正数，有些是负数。正数 Beta 值指预测的方向是正相关，而负数指相反方向。比如，“入学时的第二语言水平”的 Beta 值是正数，可解释为学生以前的第二语言水平越高，那么，在四级考试中的成绩就可能越高。相反，“母语策略”的 Beta 值是负数，就应该解释为，一个学生在学习中使用这种策略的频率越高，那么他在四级考试中的成绩就可能越低。

多元回归分析的具体操作方法是：打开 Statistics 菜单项，点击 Re-

gression，然后选中 Linear。屏幕上会出现 Linear Regression（线性回归）对话框（见图 11.10），要求你选择需要分析的自变量和因变量。选择左边的变量，点击中间箭头，所选变量就能移到右边 Dependent 和 Independent（s）下面的框里。在上面的例子里（见表 11.12），因变量是四级考试成绩，自变量有母语水平、入学时的第二语言水平以及深层学习动机等共 13 个。

第三节　分析实验数据

建立实验研究的数据文件与问卷调查的数据文件在基本程序上没有太大的区别。此外，诸如 Summarize、Correlation Analysis、Compare Means、Multiple Regression 等统计程序完全可以用来处理实验数据。两种数据处理的主要区别是，调查的数据最终可以采用 Summarize 和 Correlation Analysis 的统计程序，而实验数据的分析，除了 Summarize 和 Correlation Analysis 以外，必须涉及 Compare Means 的分析程序。为了避免重复，我们只重点介绍对实验数据分析特别重要的几个方面的内容。

1．输入数据过程中存在的问题

研究新手最常碰到的问题是，实验数据输入采用什么样的格式？我们假设，在你的实验研究中控制组有 30 个学生，实验组也有 30 个学生。那么，怎么输入这些数据呢？两组学生的数据是并列排放在一起，还是按先后顺序输入两组的数据？研究新手通常对这些问题拿不定主意。他们也搞不明白，为什么数据输入格式不对会造成某些统计程序的应用困难。

如果你的数据涉及两个或更多的群体，最简单的原则就是，按照先后顺序输入这些数据。我们可以把它叫做按顺序排列的数据格式。此外，还要弄清楚受试者是属于控制组还是实验组。表 11.13 是正确的数据文件格式，这里身份号是指派给学生的身份号码。组号是区别受试者是属于控制组还是实验组，具体说，“1”代表实验组，“2”代表控制组。独立样本 t 检验和单向方差分析都要求这种按顺序排列的数据格式，若是按照表 11.14 中平行的数据输入格式，那么，独立样本 t 检验和单向方差分析就无法处理这些数据。

ID	Group	Pretest	Posttest
01	1	79	72
02	1	64	71
03	1	86	80
04	1	72	81
…	…	…	…
31	2	71	68
32	2	63	74
33	2	80	80
34	2	60	63
…	…	…	…
61	3	68	69
62	3	80	83
63	3	76	72

表 11.13：独立样本 t 检验和单向方差分析的正确数据格式（按顺序排列的数据格式）

ID	Group	Pretest	Posttest	ID	Group	Pretest	Posttest
01	1	70	72	01	2	71	68
02	1	64	71	02	2	63	74
03	1	86	80	03	2	80	80
04	1	72	81	04	2	60	63
…	…	…	…	…	…	…	…

表 11.14：独立样本 t 检验和单向方差分析的错误数据格式

与上述情况相反，配对样本 t 检验只能分析平行格式的数据，也就是两个变量并列在一起。我们前面提到过，配对样本 t 检验是对比与同一组受试对象相关的两个变量。我们不妨再看一下第七章开始部分的例2。研究者试图调查协商互动是否比改写过的指令更能提高第二语言的理解能力。测量学生理解能力的途径是了解受试者按照老师的指令可以

将多少物品正确地放在木板上。这项研究就需要使用配对样本 t 检验的方法，而表 11.15 是这种 t 检验方法所需要的合适的数据格式。

ID（身份号）	Modified instruction（修改的指令）	Negotiated instruction（协商的指令）
01	13	16
02	9	15
03	7	10
04	11	12
...	...	

表 11.15：配对样本 t 检验的正确的数据格式（平行数据格式）

2. 实验研究中的“平均数比较”

所有实验研究都不可避免地需要使用“平均数比较”统计方法，进行独立样本 t 检验、配对样本 t 检验和单向方差分析。怎么样才能保证你所选择的统计分析程序符合要求呢？下面是帮助你选择统计分析方法的几个原则：

1）比较与两组样本相关的一个变量：使用独立样本 t 检验方法，数据格式为顺序排列格式；

2）比较与两组或更多组样本相关的一个变量：使用单向方差分析方法，数据格式为顺序排列格式；

3）比较与一组样本相关的两个变量：使用配对样本 t 检验方法，数据格式为平行数据格式。

小结

SPSS 社会科学统计软件包使用极其广泛。该软件包括 3 个主要窗口：运用、输出和数据编辑。运用窗口包含分析 SPSS 数据的所有统计程序；输出窗口展示 SPSS 分析的结果；数据编辑窗口用来建立新的数据文件或读入已储存的文件。

在问卷调查的定量数据分析之前，要给数据编码、为所有的问卷编号、输入数据、检查数据。如果问卷数据按照类别进行分类，那就需要运行个别—总体相关性这一程序，用以检查类别内部的一致性。选择什

么样的统计方法处理数据取决于研究问题。获取频数和描述性统计数据的方法是运行“Summarize”；涉及两个变量的相互关系时，运行“Correlation Analysis”；如果是比较两个或两个以上的平均值，使用“Compare Means”进行统计。具体地说，比较与同一组样本相关的两个变量时，使用配对样本 t 检验方法；比较与两组样本相关的一个变量时，使用独立样本 t 检验方法；比较与两组或更多组样本相关的同一个变量时，使用单向方差分析方法。多元回归是用来确定一组自变量在多大程度上能够预测一个因变量的变化。分析实验研究数据的步骤和程序与分析问卷调查的定量数据没有多大的区别，惟一的区别是实验研究必须使用“Compare Means”的统计方法。使用独立样本 t 检验方法和单向方差分析方法对数据格式的要求是，控制组和实验组的数据按上下顺序排列；使用配对样本 t 检验方法时，需要使用平行数据格式，即两个比较的变量分两栏平行排列。

讨论题

1. 设计一个微型研究项目回答一个研究问题，要求 30 名受试者回答一份调查问卷。
   1）建立数据文件。
   2）检查建立的概念类别是否符合统计要求。
2. 什么是相关分析？
3. 独立样本 t 检验和配对样本 t 检验之间有何区别？
4. 什么时候使用 t 检验方法，什么时候使用单向方差分析方法？
5. 使用 SPSS 提供的样本数据完成下列任务：
   1）使用相关分析计算出初始工资与现行工资之间的相关系数，现行工资与受教育水平之间的相关系数。
   2）比较男性和女性的现行工资，确定这两组职工的现行工资有无显著差异。
   3）比较女职工的初始工资和现行工资，分析现行工资与初始工资之间有无显著差异。
   4）运用单向方差分析的方法，分析在 3 组受教育水平不同的职工之间，他们的现行工资有无显著差异。
   5）使用多元回归分析的方法，找出能够预测职工现行工资的相关变量。

第十二章　定性数据分析

这一章讨论如何分析定性数据。本章分3个部分：1）分析定性数据的3个选择；2）如何准备定性分析所需要的数据；3）如何成功地进行定性分析。

第一节　分析定性数据的3个选择

第五章中提到定性数据可以用定性的方法分析，也可能需要用定量的方法分析。定性分析是指其分析结果用文字的形式表述，而定量分析的结果是以数字的形式出现。图12.1提供了分析定性数据的3个选择。

1. 定性分析

第一种选择是研究者从数据中找出各种范畴类别，这些范畴类别可能形成某些带有规律的共性或变异。例如，在分析有关使用词汇策略的访谈数据时，研究者反复阅读这些数据，最后得出了3个范畴类别：1）用来记忆生词的策略；2）用来理解词义的策略；3）用来克服词汇量不足的策略。在每个类别中，研究者又分别建立了一些小的类别。建立这些大小不同的类别是为了回答研究问题：英语专业学生在学习词汇时使用什么样的策略？

2. 定量分析

第二个选择是研究者通过计数的方法获得定性数据的统计结果。例如，在文秋芳、郭纯洁（1998）的一项研究中，受试者写一篇看图作文，研究者要求受试者在写作过程中说出自己心中的想法，并根据受试者口头记录，数出记录中出现的汉字的数量，最后参照有声思维的总字数，计算出母语词语出现的比例。统计的结果回答了研究者提出的研究问

题：第二语言写作过程中在多大程度上使用母语？

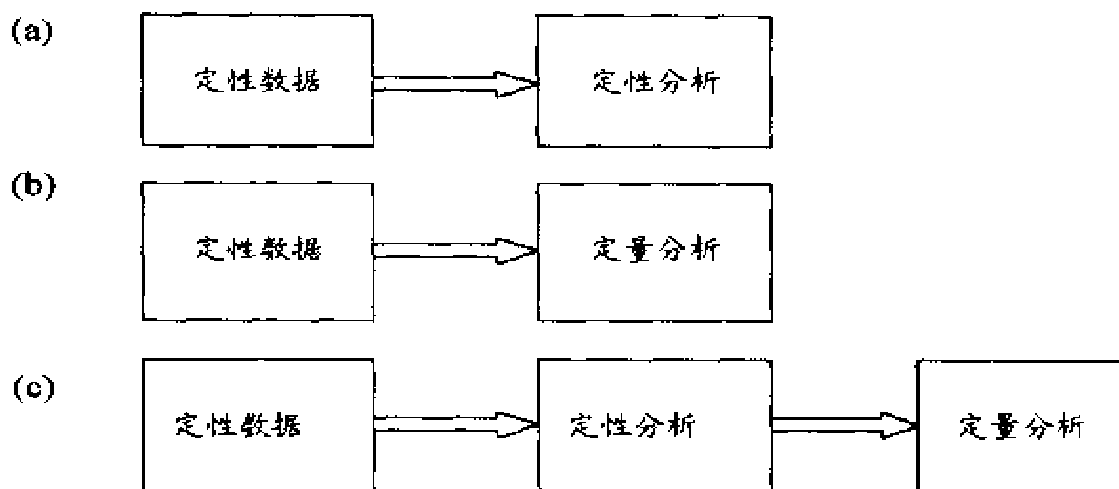


图 12.1：分析定性数据的 3 个选择

3. 先定性后定量

与前两个选择相比，这个方法要复杂一些，很容易与图 12.1 中 b 所示的方法相混淆，因为单从对研究问题回答的角度看，统计的结果都是定量的。但是这里的定量结果不是直接从原始的定性数据中获得的。我们再来看一下在第五章最后列举的例子 Chen（2000）的研究。

她的第一个研究问题是："中国孩子在记忆英语生词拼写中使用策略的频率有多高?"表 12.1 是她在这个问题上的发现。

表格中的频数结果是定量数据，但问题是这些频数在原始数据中没有。换句话说，研究者不能从有声思维记录中直接获得这些频数。研究者反复研究了受试者的口头记录，然后建立了 10 个策略类别。这个建立类别的过程往往被我们忽视。实际上，这个先建立类别，再进行频数统计的数据分析过程，必须在论文的研究方法部分详细交代。

策略范畴	频数									
	0	1	2	3	4	5	6	7	8	9+
硬记	28	5	16	16	18	8	5	1	2	1
联系形式	54	3	14	15	5	1	5		1	2
熟悉的词汇	5	1	8	8	11	13	23	6	9	16
语音方法	36	3	16	16	15	6	4	1	1	2

（见下页）

（接上页）

策略范畴	频数									
	0	1	2	3	4	5	6	7	8	9+
音节	5	1	8	13	15	12	17	9	11	9
构词方法	60	3	11	8	7	2	9			
发音规则	26	2	20	12	18	12	17	9	11	9
视觉框架	55	7	22	8	5		2	1		
拼音	73	5	15	5	2					
随意切割	44	7	21	8	4	10	3	3		

表 12.1：从有声思维数据中获得的在记忆生词拼写中使用策略的频数

上述 3 个分析定性数据的方法是相对独立的，选择哪一种方法取决于具体的研究问题。了解清楚在自己的研究中采用什么样的分析方法是很重要的，可以避免在分析定性数据和描述数据分析时产生困惑。

第二节 数据的准备

做定量分析时，研究者需要建立数据文件，需要编码、输入和检查数据。和定量分析一样，定性分析也需要建立有效的数据文件。显然，建立定性数据文件比定量数据文件的要求更高、更难。一般来说，建立定性数据文件需要完成 3 个任务：（1）转录磁带上的声音资料；（2）按照具体研究问题，分离相关文字资料；（3）用图表展示分离的文字资料。如果定性数据是日记，那么就无需做第一项任务。如果定性数据是有声思维的记录，就无法完成第二、第三项任务，因为有声思维记录的是思维的自然流露，而不是专门针对某一个研究问题所作出的反应。下面将讨论如何完成上述 3 项任务。

1. 转写磁带的声音资料

使用访谈和有声思维方法收集数据时，研究者一般将参与者的应答录在磁带上。研究者必须首先将磁带上的声音转写成文字后，才能做数

据分析。转写工作枯燥单调、耗时费力，需要研究者极大的耐心，但这是数据分析必不可少的重要部分。转写质量的高低对研究的成败影响极大。如果转写的质量差，轻则降低研究的效度和信度，重则导致整个研究失败。所以，研究者完成这项任务时必须掌握一定的专业技能，还要有强烈的责任感。

转写的首要原则是"忠实性"。研究者必须记录从磁带上所听到的一切内容。如果有听不清楚的地方，最好请受试者澄清，千万不要凭借自己的猜测和想像。第二条原则是使用标准化的转写方式。所谓标准化，是指用一套标注代码来说明停顿、重音和犹豫等现象。比如说停顿，是否要标注出停顿的长度；如果需要的话，用什么方法说明停顿的长度。在转写过程中，如果有一套这样的标注代码，就能使记录的文字标准化，日后这些数据的处理就易于操作。

2. 分离数据

数据收集的方法如果是访谈和日记，我们就可以根据具体的研究问题分离出相关数据。需要注意的是，这种数据的重新编排不能操之过急。研究者必须反复阅读原始数据，保证选择的部分确实能代表采访对象对相关问题的回答。如果急于求成的话，很可能会疏忽一些有价值的信息。此外，分离出来的数据中每个词，每句话都应该来源于原始数据。

我们以 Kwo（1999）的研究为例。在她的研究中，受试者是参加香港大学教育系为期一年的全日制教师证书课程班的学生。这个研究课题旨在调查通过全年的课程培训这些学生进步的共性规律和差异性。受试者针对课程的 6 个时段要求撰写反思笔记，这 6 个时段是：课程开始的时候、教学见习之前、教学见习之后、教育实习之前、教育实习之后、课程结束的时候。

下面介绍的分离数据文件所要回答的研究问题是"在进入该培训课程的时候，有教学经验的学员和教学新手之间在对'好教师'的看法上有何差异?"。研究者选择了 6 个学员进行深入分析，3 个有教学经验，另外 3 个没有任何教学经验。除了受试者的背景材料之外*，引用的词语、句子都是来自学员的反思笔记。他们需要回答的问题是："你对一个好教师的看法是什么?"

---

* 这 6 位学生的名字都不是真名。

第一组：有教学经历的教师

受试者1：（印度人）苏珊

母语：他加禄语　　年龄：26　　性别：女

婚姻状态：未婚　　教学经历：4年

★ 以学习者为中心

“我个人认为，这是决定教师教学效果的主要因素。对我来说，一个教学效果好的教师应该根据学生的特点和他们的需要调整自己的教学内容。”

★ 学生性格的铸造者

“我坚信，教育的主要目标是铸造学生的性格和个性。教育不能只给年轻学生灌输专业知识，而忽视人的培养，否则培养出来的学生是技能娴熟的机器人和社会能力极其低下的人。我这么说，没有不重视学生学术能力培养的意思，我只是想提醒我们的教师，教育的主要目的应该是培养学生解决日常生活问题的能力，也就是说，使他们能够对付生活中的风风雨雨。”

受试者2：凯瑟琳

母语：汉语　　年龄：30　　性别：女

婚姻状态：已婚　　教学经历：5年

★ 能够帮助学生在智力上和精神上得到发展

★ 有好的品质：谦虚、诚实、专一、不虚伪

受试者3：温蒂

母语：汉语　　年龄：42　　性别：女

婚姻状态：已婚　　教学经历：? 年

★ 求知欲强（不断更新自己的专业知识）

★ 对待学生持积极的态度

★ 对同事、学生和自己诚实

★ 富有激情（喜爱教学工作）

★ 奉献（对教学工作有强烈的责任感）

第二组：没有教学经历的教师

受试者4：希茜莉娅

母语：汉语　　年龄：22　　性别：女

婚姻状态：未婚　　教学经历：0

★ 好的个性：不发号施令、友好，乐于助人

★ 掌握好本学科的知识：教师不是百科全书，但要熟悉自己的学科，上课系统性要强，课前备课要充分。

★ 评分体系公平

★ 上课效率高

受试者 5：露西

母语：汉语　　年龄：25　　性别：女

婚姻状态：未婚　　教学经历：0

★ 耐心，爱学生

★ 讲解能力强、英语口语流利、善于与青少年交流，并能和学生建立良好的关系

受试者 6：瑞贝卡

母语：汉语　　年龄：22　　性别：女

婚姻状态：未婚　　教学经历：0

★ 有教学热情

★ 课前要备课

★ 关心学生，没有偏见

★ 不断学习，努力提高自身水平

以上是针对具体的研究问题分离出来的数据文件，有了它，我们的分析就会更容易、更方便、更深入。

3. 展示数据

数据展示是指对分离的数据文件作进一步的删除、压缩和综合，这样就可以使用表格、图表或小结的形式来表示。所谓压缩和综合，是指删除诸如引用语、解释性语言，并将一些组织松散的语句进行归纳。这种方法展现出来的数据更加简洁明了，研究者可以集中注意力去思考那些最重要数据的意义。但是压缩和综合数据也有可能会歪曲原始数据的意思。要避免这样的情况，可以采取两种办法：一是请另一个研究者检查你整理的数据。如果行不通的话，一个星期后自己重新做一次，看看第二次整理出来的数据与第一次的数据是否有差异。毫无疑问，不管出于什么原因，在对数据进行一系列处理的过程中，数据不应该被歪曲。

现在我们看看表 12.2 中的例子。表格中苏珊的回答只保留了两个短语（即以学生为中心和性格的铸造者），删除了苏珊的两大段原话。同样，希茜莉娅的回答，除了第四点与分离的数据文件相同外，其他 3 点都有很大的改动。第一点被压缩了，但还保留了希茜莉娅自己的措辞；研究者对第二点进行了概括，使用了研究者自己的语言（即知识面宽）；第三点包含了原来数据中的两个思想，研究者将其合二为一。

（第一组：有教学经历的教师）		
苏姗	凯瑟琳	温蒂
★ 以学生为中心 ★ 性格的铸造者	★ 帮助学生在智力上和精神上得到发展 ★ 谦虚、诚实、专一	★ 更新学科知识 ★ 对学生持积极的态度 ★ 诚实、富有激情、奉献
（第二组：新教师）		
希茜莉娅	露西	瑞贝卡
★ 不发号施令，友好，乐于助人 ★ 知识面宽 ★ 上课系统性强，效率高 ★ 评分体系公平	★ 耐心，爱学生 ★ 英语口语流利 ★ 善于与青少年交流 ★ 讲解能力强 ★ 热衷于教学	★ 关心学生，不持偏见 ★ 备课 ★ 不断深造，提高自己

表 12.2：数据展示：对好教师的看法

第三节　定性分析

正如本章一开始提到的那样，定性数据的分析既可能使用定性的方法，也可能使用定量的方法，或两者兼用。在第十一章，我们已经详细地讨论了定量分析，这里不再赘述。这一节只讨论定性分析。

1. 定性分析的框架

如何进行定性分析？这个问题很难用简短的语言表达清楚。如果你有足够的时间，拿出足够的耐心，和我们一起从头至尾分析某个项目的定性数据，经历这样一个完整的过程以后，相信你就会对定性数据的分析有比较深入的了解。这里，我们首先介绍由 Strauss & Corbin（1998）提出的定性分析框架，并用图形来描述这个框架（见图 12.2）。他们的框架被称为内嵌理论。这就是说，定性数据中隐含着一个理论，通过对数据的系统分析，这个隐含的理论逐渐从数据中显露出来。接着，我们将以一个简单的实例说明如何运用这个框架。需要说明的是，这里的介

绍比较简单，对做硕士论文或博士论文还远远不够。除了阅读本小节的内容外，我们向大家推荐两本书：一本是由 M. B. Miles 和 A. M. Huberman 撰写的《定性数据分析》（Qualitative Data Analysis，1994），另一本是 A. Strauss 和 J. Corbin 撰写的《定性研究入门》（Basics of Qualitative Research，1998）。

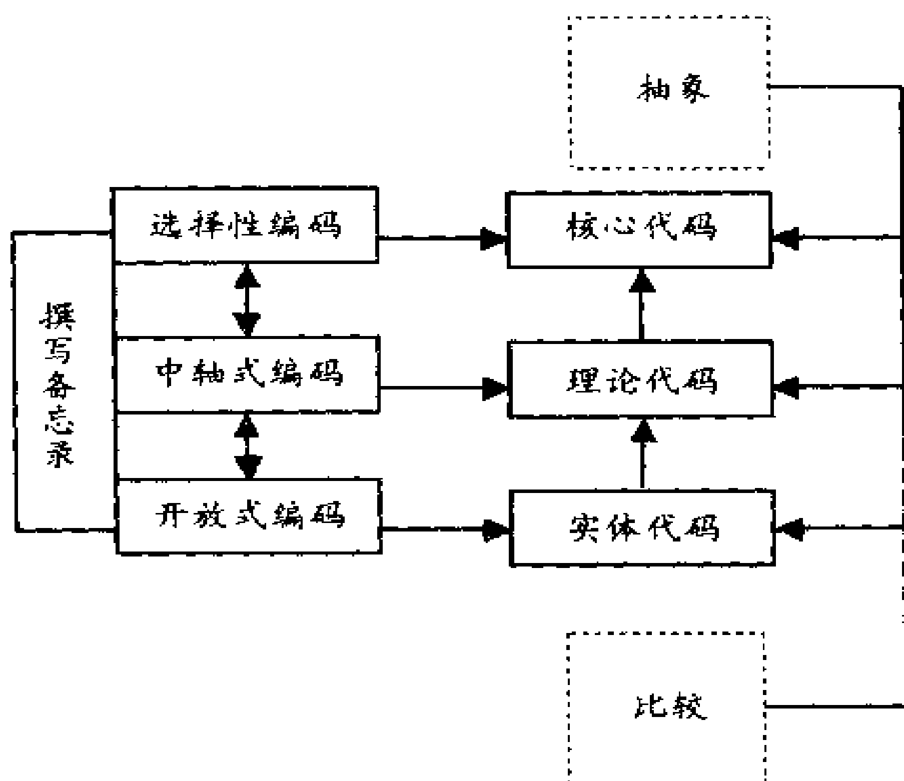


图 12.2：定性数据分析框架

根据图 12.2，定性数据分析包含两个基本任务：编码和撰写备忘录。编码又分为 3 种：开放式编码、中轴式编码和选择性编码。这 3 种编码方法又分别产生 3 种代码。所产生的代码按照抽象的程度形成了一种等级体系，也就是说，由开放式编码方法产生的实体代码处在最低层，由选择性编码方法产生的核心代码处在最高层，由中轴式编码产生的理论代码处于两者之间的中间层。撰写备忘录的目的是记录研究者在编码过程中的想法。所作的记录是多面的，可能是研究者对数据的分析、解释，也有可能是研究者对数据的评价。总之，这些记录是在数据编码的过程中研究者所有的内心活动。

完成编码和撰写备忘录这两项任务主要涉及两类心智活动：抽象和

比较。所谓抽象，就是寻求一个概括性程度比较高的陈述，用来描绘和解释某种现象或概括性不强的概念。所谓比较旨在寻找物体之间或概念之间的相同之处和不同之处。

2. 心智活动的基本操作

在这一小节里，我们将解释抽象和比较两类心智活动的基本操作，以及描述完成编码和撰写备忘录两项任务的基本程序。

抽象和比较是定性数据分析中的最基本的心智活动。也就是说，不管你做哪种编码，你的心智活动都离不开抽象和比较。

2.1 抽象

"抽象"与"具体"相对，但抽象的程度不等，可以是比较抽象，也可以是非常抽象。编码产生的结果可以按照抽象程度的高低，放在不同层次上（见图 12.3）。抽象程度的最低层次是概念，表示一个具有共性或表现出相似属性的单位。上一个层次是范畴，表示一组具有相同性质的概念。假设又比范畴更加抽象。假设是指研究者对范畴与次范畴之间关系的解释。抽象程度最高的是理论，表示一组在同样范式里的假设相互关联。

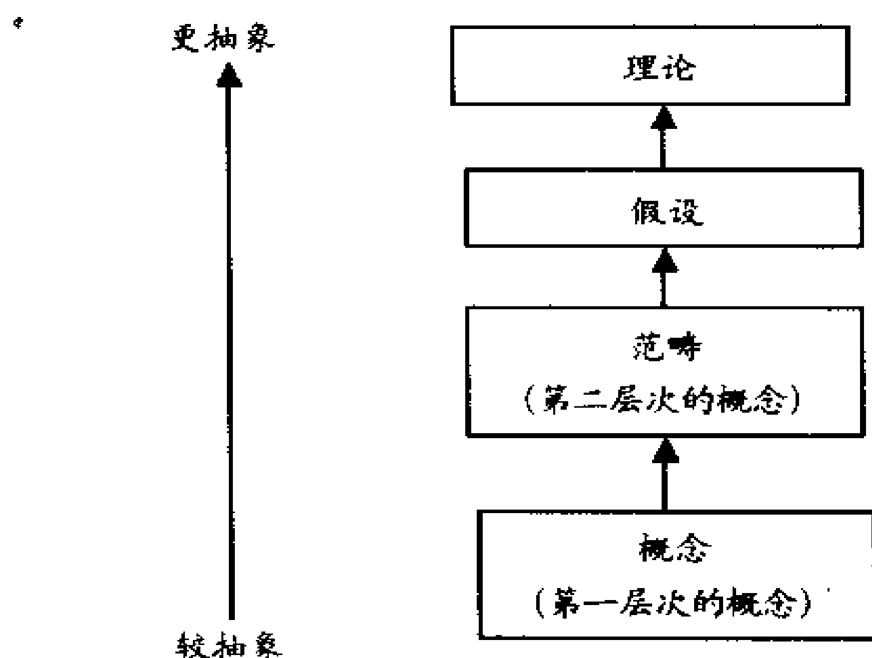


图 12.3：抽象等级层次

就单个的研究项目而言，它的目的不一定是为了建构一种理论。确定哪一个抽象层次作为研究目标取决于你的研究问题。不管想要达到什么样的抽象层次，一定要先确定一个共性的东西。这个共性的东西可能是一种特性，一种属性，或是一种范式。对于较低层次的抽象，我们寻找的是各个项目共有的特性或相似的属性。例如，“液体”这个词可以指石油、玉米油或水。它们尽管在颜色、质量和功能等各个方面有所不同，但都有一个共同的特性：能够流动。对于较高层次的抽象，我们感兴趣的是一种范式，各个范畴之间的相互联系依靠的就是这种范式。例如，液体、气体和固体都属于“没有生命的材料”这个范式，因此统称为“物质”。通常，研究者，特别是研究新手，不容易找到共同的特性和范式。

从抽象的程度来看，编码的结果一目了然，但是我们大脑的活动可没有这么简单。也就是说，我们大脑的活动程序不是按照由低到高的抽象层次，循序渐进。大脑通常是同时处理概念、范畴、假设和理论，各个层次之间彼此还交叉互动。有时候，我们从理论到概念，有时候是从概念到理论。总之，我们思维的路径多种多样。但是，为了便于叙述和讨论，我们将思维过程分解开来，目的是说明 3 个层次的编码程序与 3 个层次的抽象程度互相对应：开放式编码产生概念和范畴，中轴式编码产生假设，选择性编码产生理论。我们希望，这种分解方法不会给大家造成错觉，误认为定性数据分析可以按照一组固定程序进行。其实，定性分析最难的就是没有固定的程序。

2.2　比较

“比较”是数据分析中又一个基本的心智活动。它与“抽象”活动相互渗透，相互交融。“比较”可以出现在不同抽象等级的各个层次上。在较低的抽象层次上，可以比较事件、思想或行为，并把这些类似的事件、思想或行动归入某个概念。确定了概念之后，就需要将不同的概念进行比较，以确定它们之间的共性与差异，最终建立相应的范畴。同样，范畴之间的比较可以帮助我们看出它们之间的关系。正是通过这样的比较，我们才能找出共性的东西或范式。“比较”不仅可以在某个研究内部进行，也可以在两个或几个研究之间进行。

3. 两个基本任务

3.1 编码

从理论上说，开放式编码、中轴式编码和选择性编码有着本质性的差异。开放式编码的目的是识别实体代码，中轴式编码的目的是识别理论代码，而选择性编码的目的是识别核心代码。虽然它们的操作方式各不相同，但没有必要按照顺序逐一完成。在许多情况下，它们是彼此交叉的，可以同时进行操作。

开放式编码 开放式编码过程要求研究者考虑到数据中的各种可能性。也就是说，研究者不能定向思考，不能局限于一种选择。开放式编码最好不要一个人来做，而是由一个研究小组或群体来做。如果行不通，可以先将这些临时的代码搁在一边，过几天再来分析这些数据，这个过程可以反复多次。这样做的好处是，研究者能在数据中寻找出各种理论的可能性。

开放式编码首先要发现数据中的概念，然后对确认的概念进行归类。整个数据中的每个事件、每个行为或每个思想都要仔细地观察。通过对事件与事件的比较，找出它们之间的相同之处和不同之处，在这个基础上，再确立概念和概念的归类。通过这种方法得出的最终概念或范畴在内嵌理论中被称为实体代码。

建立范畴非常重要。它们不仅能够减少分析单位的总数，而且便于记忆和思考。更重要的是，这些范畴可以分解成次范畴，有助于我们做更深入、细致的分析。范畴的取名通常取决于研究的视角，因为同样的现象可以有不同的名称。范畴的取名应该有一定的逻辑性和透明性。所谓逻辑性、透明性，是指取名要有一定的道理，所指明了易懂。

我们以 Wen & Guo（1998）的定性研究为例。该研究运用有声思维的方法调查了母语思维在英语看图作文中的作用。结果表明，母语思维的主要功能包括转换中介、内容验证中介、内容生成中介、二语语言形式检索中介和写作程序管理中介等 5 个功能。这 5 个功能的名称实质上就是范畴，其中有些被分成了更细的次范畴。例如，“内容生成中介范畴”包含下列次范畴：推理、判断、联想、提问、监控和评价。

编码的整个过程不是一个线性过程：从概念到范畴再到次范畴。实际上，一开始抽象出来的概念只是临时性的，需要不断修正，必须保证

生成的每个概念或范畴能够解释所有的数据。如果出现自相矛盾的地方，就应该修改初步形成的概念或范畴，这样的修改过程不可能一次性完成。从本质上讲，范畴的形成是循序渐进的。但是，为了避免过于频繁的修改，最好在确立概念或范畴之前，多花一些时间去认真研究数据。

中轴式编码 开放式编码是将数据分解成不同部分，并考虑到数据中存在的各种理论上的可能性。与开放式编码不同，中轴式编码对各个范畴进行重新组合，通过寻找连接各范畴以及它们的次范畴之间的中轴，从而形成一个完整的画面。建立各范畴之间的联系，可以帮助我们对有关现象作出更准确和更有力的解释。尽管这种联系是在中轴式编码过程中提炼归纳而成，但理解各个范畴彼此之间相互关系，在开放式编码过程中就已经开始。从这个意义上说，开放式编码和中轴式编码之间并没有明显的界限。我们再来看一下“内容生成中介”这个范畴。我们并不是按照先确立内容生成中介范畴然后寻找相关次范畴这个顺序，也不是相反的顺序。实际上，我们在看数据的过程中，范畴和次范畴几乎是同时形成的。

找出各个范畴之间联系的中轴是中轴式编码的关键。但什么是中轴呢？简单地说，这是在内嵌理论中，观察数据的一种视角或范式。它也许是某种现象发生的条件，这些条件涉及时间、地点、方式和原因，构成这种现象发生的情景结构。它也有可能是在特定条件下随着时间而发生的一系列行为，这些行为表示着某现象发生的一个过程。中轴也可能是这些行为的结果。我们还是以内容生成中介这个范畴为例。该范畴包含推理、判断、联想、提问、监控和评价 6 个次范畴。显然，这 6 个不同的行为在心理上有着本质的不同，但它们都是写作过程中生成写作内容的行为。因此，在这个例子中的“中轴”是“功能”。换句话说，这 6 个活动都是服务于同一个功能、生成写作的内容。

Strauss（1987）指出，中轴式编码包括 3 个主要活动：1）说明范畴的特性以及该范畴的各个维度；2）确定与某个现象相关的各种可能的条件、行为和结果；3）寻找数据中能够确立范畴之间关系的线索。

Strauss 和 Corbin 强调，寻找中轴或范式只是手段，而不是目的。真正的目的是寻求对现象的解释和深入理解，而不是寻求某个现象的条件、行为和结果。

选择性编码 选择性编码，顾名思义，是指选择一个方面作为核心范畴或中心范畴，这个范畴代表着研究的主题。一旦选择了核心范畴，分析的范围就缩小了，分析的重点也凸现出来了。围绕这个核心范畴，我们

可以聚集所有的相关范畴，在此基础上形成一种理论。很显然，核心范畴是构建一种新理论的关键　那么，什么样的范畴才算是核心范畴呢？如何确立这样的范畴？Strauss（1987：36）列出了衡量核心范畴的一系列标准：

1）必须是中心范畴，即所有其他的主要范畴都可能与之相关。

2）必须在数据中频繁出现，即在所有或几乎所有案例中，都能找到代表这个概念的种种指标。

3）涉及这些范畴的解释必须具有逻辑性和一致性。对数据的解释不应有半点的牵强附会。

4）用来描述中心范畴的名称或词语应该是抽象的，可以用于其他直接相关的研究领域，最终能形成更抽象的理论。

5）由于概念是通过与其他概念的整合分析提炼出来的，因此，形成的理论具有深度和解释力。

6）概念除了解释数据中的要点外，还要能解释各种变化。也就是说，无论条件发生什么变化，或是表述某个现象的方式有什么不同，核心范畴的解释力保持不变。此外，核心范畴概念还应该能够解释矛盾的案例或备选案例。

但是只了解这些标准还不够，还需要勤于实践。有些研究者，特别是新手，虽然熟悉这些标准，但就是找不出核心范畴，这种情况很常见。他们面对浩如烟海的数据不知所措，更不知道如何从不同的角度去看这些数据。Strauss & Corbin（1998）针对如何选择核心范畴，如何整合概念，对研究新手提出了几条有价值的建议。第一条建议是撰写“故事梗概”，也就是写上几句话，描述该研究的大概。开始的时候可能会觉得很难入手，这时候可以回过头来看原始数据，重读几个案例，并不断问自己，“这里的主要问题是什么？”，“最引起我注意的是什么？”。第二个建议是使用图表。图表是显现各个概念之间关系的非常有用的工具。画图表的时候，对概念之间的逻辑关系必须有清晰的思路，否则，画出来的图表反而使人感到迷惑。第三条建议是回顾和整理备忘录（如何撰写备忘录将在下面讨论）。备忘录是用来记录在编码过程中大脑中闪现的想法。随着数据分析的深入，我们的备忘录会包含越来越多的抽象内容，这对我们抽象出核心范畴很有帮助。

3.2 撰写备忘录

撰写备忘录是定性数据分析的一项基本任务，目的是记录通过抽象和比较所获得的信息以及编码过程中大脑里闪现的灵感。撰写备忘录会迫使你特别注意数据，提醒你思考的方向。有了备忘录，你的分析结果在随后的分析过程中很容易得到修改、补充、合成或删除。正如 Glaser 在其所下的定义中阐述的那样，撰写备忘录是一种非常灵活的方法：

> 备忘录是数据分析者在编码过程中对代码及代码之间关系的思考的理论概括……它可以是一个句子，也可以是一段话，甚至是几页的长篇叙述……它穷尽了数据分析者在不同阶段对数据形成的各种想法，或许还带有少量的理论阐释。（Glaser，1978：83－84）

撰写备忘录与编码应该同时进行，从编码过程中获得的任何结果都应该写入备忘录。备忘录可以记载有关实体代码的想法，也可以记载有关理论代码或是数据分析方法的看法，也可以记载对有关个案的看法。备忘录如果是有关实体代码和理论代码的笔记，那么备忘录的主要内容就应该与概念有关，必须超越分析的内容。换句话说，备忘录应该是从实际的事件到抽象的概念，从低层次的抽象到高层次的抽象。

有一点应该记住，备忘录是随着时间变化的，没有对错好坏之分。实际上，备忘录的内容具有临时性的特点，随着分析的深入，其内容会发生变化。回头看看最初的笔记，你可能会觉得幼稚可笑。但是，随着数据分析的深入，备忘录会变得越来越复杂，范畴、模式或关系就会从数据中慢慢凸现出来。只要不断记录分析过程中产生的想法，不断根据新的数据修改这些想法，你的分析就会进入越来越高级的抽象层次，最终，你的备忘录就会成为建立一种理论的主要依据。

作为实例说明，我们再来仔细观察表 12.2 中的数据，并进行定性分析。研究问题是，进入该培训课程之初，有教学经验的学员和教学新手在对“好教师”的看法上有何差异？

对这个问题，我们应从哪儿入手？实际上没有固定的程序。这里我们只能简单地描述研究者分析数据的程序。我们知道，人的思维不是单向的，或线性的，在我们的大脑里，通常好几个想法会同时出现。然而，我们需要用语言来逐一描述，给人的感觉好像这些想法的出现是遵照了一定顺序。从这个意义上讲，下面研究者的自述不是数据分析过程

的真实描述。

分别找出第一组和第二组中的共同点

阅读表中第一行数据的时候（即第一组），“性格的铸造者”和“精神上的”两个词引起了我的注意。我在备忘录中写道：有教学经历的老师似乎极大地关注学生的心理发展。但是第3个受试者温蒂没有使用类似的词语。通过对凯瑟琳和温蒂的比较，我发现她们都提到了一个好教师应该具备的一些素质，并将这个想法记在备忘录里。到了这儿，我被卡住了，我看不出3个受试者之间的看法有什么相同之处。我没有绞尽脑汁再去思考第一组受试者，而是开始观察第二组的数据。读了第一遍，我马上就发现了第二组数据中的两个共同特征。第一个共同特征是受试者都注重教师与学生之间的关系；第二个特征是她们都强调教学的效果。于是，我制作了一张表格，说明这两个共同特征和相关证据。

姓名	与学生的关系	有效的教学
希茜莉娅	不发号施令，友好，乐于助人； 评分体系公平	知识面宽，上课系统性强，效率高
露西	爱学生；善于与青少年交流技巧	口语流利；讲解能力强
瑞贝卡	关心学生，不持偏见	热衷于教学；备课；不断学习，提高自己

表 12.3：3个没有教学经验教师的两个共同特征

在第二组中找到了两个概念或实体代码之后，我又回到第一组检查该组是否也有这两个特征。有趣的是，没有教学经验的教师所关心的内容根本不是有教学经验的教师所注意的内容。这两组教师完全是从不同的角度对待相同的问题。他们看待这个问题的视角有什么不同呢？我一遍又一遍地阅读数据，想搞清她们看待这个问题的差异。突然有两个词语引起了我的注意：“公平的评分体系”和“不持偏见”。这两个短语是学生常常用来谈论自己的老师，没有教学经验的教师自然会从学生

的角度看待这个问题。按照这个思路，我意识到第二个特征“有效的教学”既是学生的角度也是授课教师的角度。

分别找出两组之间的差异

脑子里有了这两个范畴，我又仔细地看了第一组的数据。我似乎觉得，在苏珊和凯瑟琳看来，一个好教师应该是教育家，应该把学生培养成一个完整的人，即在智力上和心理上都得到发展的人。温蒂的观念似乎与她们不同。她似乎是从一般教师的角度看待这个问题的，她关心的是一个好教师应具备的一般品质。到了这个阶段，我在备忘录里画了一个图表，试图按照一定层次列出思考这个问题的所有不同角度。另外，我还试图将实体代码和理论代码联系起来。

“有教学经验的教师和教学新手在对好教师的看法上有何差异?”从图表上看，这个研究问题的答案一目了然，似乎原本就不言自明。

上述定性分析的案例说明虽然简单，却清楚地向我们展示了比较和抽象的操作程序，获取实体代码和理论代码的方法，以及如何使用备忘录帮助定性数据的分析。

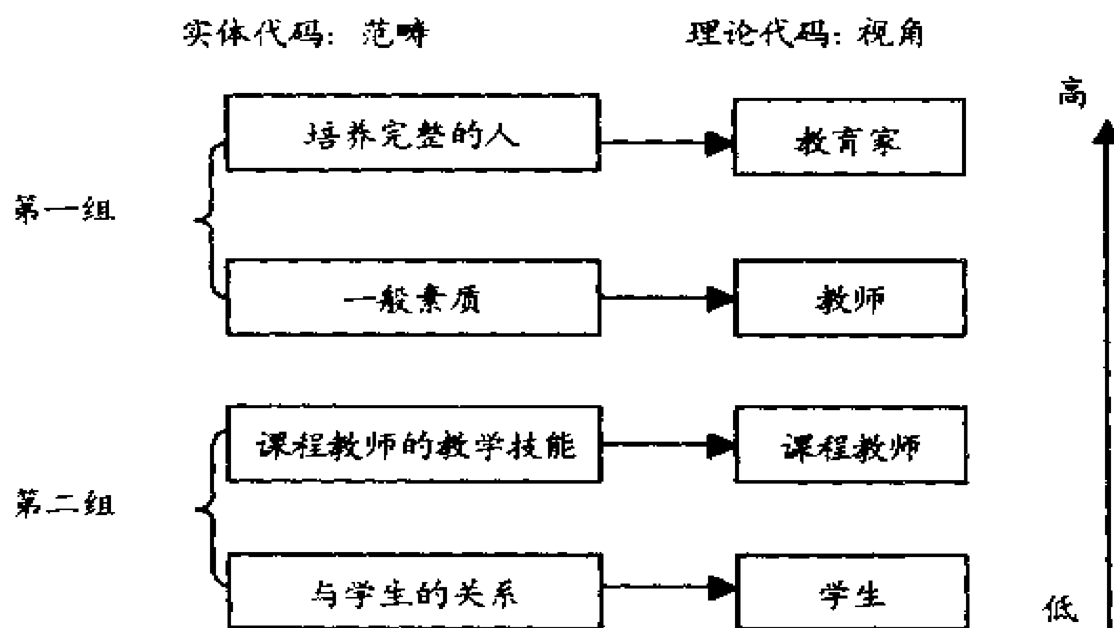


图 12.4：第一组和第二组在“好教师”看法上的差异

这里，有一点需要强调，如果研究者采取不同的研究角度，对这个问题的回答有可能不同。这也是定性分析最具有吸引力的地方。如，上

述问题可能会出现这样的答案：有教学经验的教师和没有教学经验的教师在“好教师”看法上的根本差异在于，前者关心的是职业上的发展，而后者关心的是教师的生存技能。

第四节 效度和信度

定性分析常遭到的批评是缺乏效度和信度。如果没有采取有效的措施防止数据分析中的缺陷，效度和信度的确会受到影响。我们在前面说过，定性分析没有固定的程序帮助我们得出可信的结果。从数据的表面，找不出实体代码、理论代码和核心代码，研究者只有通过反复阅读数据，不断思考，才能抽象出这些代码。

定性分析的有效性，是指所发现的代码能够解释数据。换言之，代码应该完全受数据的支持，研究者不能将自己的想法强加于数据。定性分析的可靠性，是指两个或两个以上的研究者针对同样的一组数据能够获得类似的代码。

要保证定性分析的效度和信度，两个研究者最好独立地分析相同的一组数据，然后一起讨论分析的结果。如果有些争议不能解决，应该请第三方加入讨论的行列。

小结

定性数据可以用定性的方法或定量的方法分析，也可以两种方法一并使用。如果要分析访谈和有声思维数据，就必须将磁带上的声音转写成文字，转写时必须遵守“忠实性”的原则，还应有一套标准化的标注符号用来说明停顿、重音和犹豫等现象。如果磁带有听不清楚的地方，最好请受试者来澄清，不要凭借研究者自己的主观猜测和想像。研究者最好根据具体的研究问题对定性数据进行分离，然后通过删减和合成展示分离的数据。分离和展示数据对定性分析极为有利。按照 Strauss 和 Corbin 提出的分析框架，我们应该完成两个最基本的任务：编码和撰写备忘录。通过“比较”和“抽象”这两个主要的心智活动，研究者首先要做开放式编码，接着是中轴式编码，最后是选择性编码。这 3 种编码方式又产生 3 种代码：实体代码、理论代码和核心代码，这 3 种代码可以放在由低到高的不同抽象层次上。必须注意的是，这 3 种不同的编码

方式不是进行数据分析的固定程序，整个编码过程是一个使用直觉的过程，也是一个互动的过程。撰写备忘录时，应记下比较和抽象过程中所有的心理活动。你心里的想法最初可能是零碎的或杂乱无序的，但至少可以说明分析数据的进展过程，也可以作为获取洞察力的基础。为了提高定性分析的效度和信度，最好请另一个研究者分析相同的一组数据，也可以经过一段时间间隔之后由研究者本人重新分析相同的数据。

讨论题

1. 到图书馆查找硕士生毕业论文，以这些论文为例，说明如何用定性或定量的分析方法或两种方法并用来分析定性数据。
2. 对磁带录音进行文字转写时需要注意哪些方面?
3. 在对数据进行分离时，应该采取哪些方法?
4. 在展示数据时，如何处理分离的数据文件?
5. 如何区分3种不同的编码方式?
6. 备忘录的作用是什么?
7. 如何保证定性数据分析的效度和信度?

第三部分 论文写作

论文写作的成败决定着研究生能否顺利毕业。虽然论文写作的重要性人所共知，但研究生经常在这方面得不到系统的训练和指导。有人经过艰苦努力，最后过关了；也有人费了九牛二虎之力，仍旧失败了。这里我们将和读者分享过去成功的经验和失败的教训，以期能从中获益。这一部分共分 3 章：论文的结构、论文的撰写、论文的格式与风格。通过这 3 章的学习，要能做到：

- 熟悉论文的结构；
- 掌握论文每部分的内容与撰写程序；
- 解决论文写作格式中的一些典型问题。

第十三章　概　述

本章首先讨论论文的界定，接下来介绍论文的总体结构，最后对论文的每个部分进行简要描述。

第一节　论文是 thesis 还是 dissertation?

关于论文是 thesis 还是 dissertation，目前尚没有一致的看法。Cone & Forster（1993）通过比较《韦氏国际大词典》与《兰登英语大词典》，得出的结论是 thesis 和 dissertation 之间没有明显的区别，但在美国大学里，thesis 多指硕士论文，而 dissertation 多指博士论文。根据笔者在英国教育制度中的经历，thesis 这个词既可以指硕士论文，也可以指博士论文。

什么是论文？论文是讨论和研究某个问题的文章，它的各部分在逻辑上要相互联系，且一切都要围绕研究问题（Punch，1998）。论文既不是单纯记录研究过程中的所作所为，也不是简单描写研究中的发现。论文是对某个问题进行论辩，因此除了描述以外，还必须要论证研究设计的合理性。

研究过程或多或少都有点紊乱，这一点不足为怪。但要使紊乱的研究过程在论文中显得井井有条、详略得当，研究者必须认真审视整个研究过程，精心选择只与研究问题相关的、重要的信息进行报告。撰写论文的过程实际上是对研究结果进行选择与重组的过程。简单说来，与研究问题相关且重要的结果就报告；不相关、不重要的就置之不理。当然，这里的选择决不意味着掩盖与研究假设不相符的结果。

第二节　论文结构

图 13.1 描述了论文的总体结构，与论文的常规顺序相同。整个论文可分为 3 个部分：开篇部分、主体部分和结尾部分。开篇部分包括论文

题目页、导师签名页、英文摘要、中文摘要、致谢、目录、表格索引和插图索引；主体部分由引言、文献回顾、研究方法、研究结果与讨论和结论组成；结尾部分包括参考文献和附录。需要说明的是，论文的具体写作，并不是按照以上顺序。一般先写主体部分，接着写结尾部分，最后写开篇部分。为了叙述方便，下面的讨论仍旧按照论文结构的顺序进行。

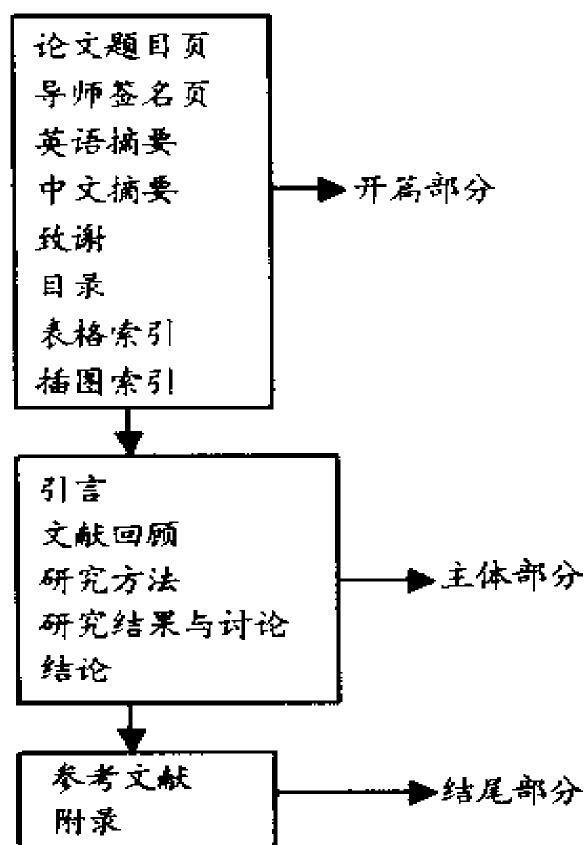


图 13.1：论文结构

1. 开篇部分

论文的开篇部分包括论文题目页、导师签名页、英文摘要、中文摘要、致谢、目录、表格索引和插图索引。下面我们对其各个部分一一阐述，并指出写作过程中可能会出现的问题。

1.1 论文题目页

论文题目页是论文的第一页，这一页包括论文题目、论文类别（是硕士论文还是博士论文）、论文作者所在学校、论文作者姓名以及完成时间。各个大学对论文格式要求不尽相同，建议你事先到研究生主管部门把论文格式了解清楚。

论文题目既要简明扼要，又要能说明问题。如果研究的变量不多，论文题目可包含这些变量。如果变量较多，论文题目可表明自变量和因变量的类型。如果有可能，最好把研究对象也放进论文题目中去。下面是一些论文题目：

1）英汉恭维语回答的对比研究（Gong，1998）
2）对中国高水平英语学习者的研究：学习者可控因素与学习结果之间的关系（Wen，1993）
3）中文语境下母语对二语写作认知过程的影响（Guo，1997）

4）中国大学生英语学习动机的内部结构研究（Qin，1998）
5）英语专业二年级学生语用能力研究（Wu，1998）
6）大学英语专业学生二语学习观念变化的研究（Su，1996）
7）大学生英语议论文写作研究：修辞知识与语篇模式（Wu，J.，1998）
8）中国大学英语课堂教学中对错误的处理（Hu，1998）
9）任务类型、准备时间长短、学习时间长短对大学生口语成绩的影响（Zhu，1997）
10）中国大学生的二语水平与理解策略的使用（Lu，1997）
11）中国大学英语专业学生冒险观念与行为研究（Wang，1999）
12）中国非英语专业大学生二语语言变量与写作能力的关系（Ma，1998）

有些大学对论文题目的字数有限制。论文题目的确定时间也有规定，例如香港大学教育学院就明确规定，论文题目必须在提交论文前3个月，上报学位委员会。一旦确定，就不能再更改。

1.2 导师签名页

导师签名页是论文的第二页。尽管这一页做起来比较容易，但这一页的信息很重要。没有导师的签名，论文就无法上缴。

1.3 摘要

摘要是研究的小结。摘要本身要连贯，自成一体。它必须包括以下几个方面的信息：研究的目的、研究问题、研究对象、收集数据使用的工具、收集和分析数据的程序、研究结果以及结论。撰写前，最好先去阅读几篇权威杂志上的摘要，分析这些摘要的结构。

在中国大学完成论文，除了中文摘要外，还需要撰写英文摘要。毫无疑问，中外文摘要的内容应该相同。但切忌把中文摘要逐字逐句翻译成英文。带有很明显翻译痕迹的英文摘要，句子读起来往往不通顺，有些甚至读不懂。为了确保可读性，写好摘要的初稿，最好请几个外系的学生仔细阅读并提出修改意见。这里需要强调的是，外文摘要和中文摘要同等重要。对于外籍专家学者来说，他们主要是通过外文摘要来了解该论文所研究的问题及其学术价值。从这个意义说，外文摘要的质量会直接影响人们对论文的整体评价。

1.4 致谢

致谢页是研究者表达谢意的地方。需要致谢的人包括所有在研究和

论文写作过程中提供过帮助的人。其中有导师、曾经提过意见或建议的老师和同学、帮助收集数据和校对的人、给予物质和精神支持的家人及亲朋好友、给予经济上资助的人（如果有的话）、允许使用他们的研究工具和资料的人。

这一部分常见的问题是致谢比较笼统抽象。如：没有提到别人的帮助体现在何处，这样的感谢貌似礼貌，但显得不真诚。另一个常见问题是过分谦虚，导致别人误认为作者能力不强。最好的办法莫过于致谢时明确说出何人在何时何处提供了何种帮助。最糟糕的是，有极个别学生抄袭别人的致谢辞，或稍作修改。这样的致谢词不仅不能起到致谢作用，相反会引起曾提供帮助人的不愉快。总之，表达谢意要诚恳，措辞要实事求是、恰如其分。

1.5 目录

目录是把开篇部分、主体部分及结尾部分的标题列出来。开篇部分页码用罗马数字标出，其他部分用阿拉伯数字标出。就论文的主体部分而言，还要列出每章的小标题，以体现其层次性。目录中不要把所有的小标题都列出来。一般列出 3 个层次的标题就行了。从表 13.1 的目录中我们可见一斑。

目录

致谢	i
摘要	iv
表格索引	xiv
插图索引	xix
附录索引	xx
引言	1
1 此项研究的必要性	1
1.1 有助于提高英语学习的效率	2
1.2 有助于推广对成功教师的新理念	4
1.3 有助于解决有关英语学习的自相矛盾陈述	6
1.4 有助于研究学习策略对二语成绩的影响	7
框架结构	9
第一部分 背景介绍	
第一章 对影响学习结果诸多因素的研究	11

（见下页）

（接上页）

1.1	引言	11
1.2	教育评估理论发展的总趋势	11
1.3	影响学习结果的诸多因素	13
1.4	学习者可控因素	15
	1.4.1 学习路径（动机+策略）	16
	1.4.2 元学习	17
1.5	教育领域中的研究方法	20
	1.5.1 两个相对的研究方法：定量研究与定性研究	20
	1.5.2 有关定量分析变量之间关系的进展	23
1.6	小结	23

表 13.1：目录举例（Wen，1993）

1.6 表格索引

表格索引集中了论文中所有的表格编号和其所在的页码，这样便于读者可以很快找到所需信息。

1.7 插图索引

和表格索引一样，插图索引集中了论文中所有的插图编号和其所在的页码，这样便于读者可以很快找到所需插图信息。表格和插图必须有适当的名称，这样读者可一目了然。

2. 主体部分

论文的主体部分常见的结构模式有两种：单层次模式和双层次模式。

一般说来，论文的主体部分分为 5 章：第一章引言；第二章文献回顾；第三章研究方法；第四章研究结果与讨论；第五章结论。这是单层次模式。如果采用双层次模式，整个论文就被分成几个部分，每个部分又分为几章（参见表 13.2）。博士论文一般用双层次模式。本书只介绍单层次模式。

第一部分　背景介绍
第一章　引言
第二章　文献回顾
第二部分　研究方法
第三章　定量设计
第四章　定性设计
第三部分　研究结果与讨论
第五章……
第六章……
第四部分　结论

表 13.2：双层次模式举例

2.1　引言

引言是整个研究的一个概貌。通常较短，没有技术细节，一般读者都能读懂。写这一章的一个难点是着墨不多，却要给读者提供足够的信息。

2.2　文献回顾

文献回顾这一章要系统、全面地回顾现有的相关文献，表明前人已做过哪些研究，还有哪些问题有待解决，让读者跟随你一道形象完整地体验相关研究的情景，了解此项研究与以前研究的联系，以及此项研究对现有文献的可能贡献。

2.3　研究方法

这一章要描述研究设计。它包括这样 5 个方面的内容：1）研究问题或研究假设；2）研究对象；3）研究工具；4）数据收集程序；5）数据分析程序。这一部分的信息应该清楚明白，以便其他研究人员仿效其步骤与程序，进行重复研究。

2.4　研究结果与讨论

“研究结果与讨论”这一章包括结果与讨论两个部分。结果部分必须要回答前文中列出的所有研究问题。讨论部分要就每个研究结果提出

可能的解释，同时指出研究结果的意义，以及本次研究的结果与以前研究结果的联系。讨论既可以与报告研究结果同步进行，也可以先报告研究结果，再进行讨论。

2.5 结论

结论这一章包括主要研究结果、这些研究结果的启示以及对未来研究的建议。

论文主体部分的格式各个大学的要求可能不完全相同。撰写之前，最好向自己的导师或负责研究生的秘书了解清楚，以免走弯路。

3. 结尾部分

结尾部分包括引用书目和附录。这两者都很重要，可又不宜把它们放在论文主体部分，否则会影响论文的流畅性。

3.1 引用文献

论文写作过程中需要阅读大量著作和论文，然而，并非所有的阅读资料都在论文中引用。引用文献，顾名思义，这些文献一定是论文中引用过的资料。按照美国心理学协会（APA）引用文献的格式，论文中所有引用的资料按照作者的字母顺序排列。每个引用文献条目应包含以下信息：

1）作者；

2）出版日期；

3）题目；

4）如果是著作，必须标明出版社和出版地；

5）如果是论文，必须标明杂志名称、卷数、期数和页码范围。

研究生撰写"引用文献"这一部分时，常见问题有两个。一是将 Bibliography 和 References 混为一谈。在中文译文中，这两个英文词都可以译为参考文献。但在英文中这两个词有区别。Bibliography 可包括研究中参考的任何资料。即使没有引用，也可以列进去。而 References 只包括在论文中所引用的资料。我们建议为了避免混淆，最好把 References 译成"引用文献"。根据美国心理学协会的规定，论文要求列出的是 References。第二个问题是，引用文献中包括了一些没有引用的文献，而有些引用的文献又没有被列进去。出现这个问题的主要原因是由于反复修改论文，引用文献没有及时修改。如何避免这些问题？可以把在写论

文期间查阅的相关文献列个清单，等论文完全修改结束后，对照论文把清单中被引用的文献一一挑选出来，然后放到引用文献中去。答辩委员会成员经常指出引用文献中存在的问题，例如论文中所引用的文献在引用文献中没有列出，或者列出的引文文献条目格式不统一。

如果在最后一刻，仍搞不清作者的名字如何拼写或文献准确的出版日期，最好放弃这一文献，重写这一段内容。否则在浩如烟海的资料中查找某篇文献，耗时费力，得不偿失。

3.2 附录

附录部分可包含作者认为重要的，可又不能放在论文正文中的任何材料。常见的附录包括以下内容：

1）完整的问卷；
2）采访计划与安排；
3）观察提纲；
4）考试试题；
5）培训中所用的材料；
6）先导测试的研究报告；
7）要求研究对象所承担的任务，诸如读、写、听、说等；
8）数据分析的详细结果。

小结

论文属于论述文，各部分必须存在密切的逻辑联系，所有的论述都要围绕研究问题。它由3部分组成：开篇部分、主体部分和结尾部分。开篇部分包括论文题目页、导师签名页、英文摘要、中文摘要、致谢、目录、表格索引和插图索引。主体部分包括引言、文献回顾、研究方法、研究结果与讨论、结论。结尾部分包括参考文献和附录。

讨论题

到图书馆找一本硕士或博士论文。仔细阅读开始部分和结尾部分，然后回答下列问题：

1. 这篇论文的总体结构是什么？
2. 中英文摘要中有没有问题？

3. 致谢部分写得得体吗？有什么改进意见？

4. 目录是如何编排的？有什么改进建议？

5. 表格索引与插图索引有没有什么问题？

6. 简评论文后的参考文献和附录。

第十四章 论文写作

本章的重点是论文主体部分的写作。具体地说，本章就引言、文献回顾、研究方法、研究结果和讨论以及结论各部分的写作方法逐一介绍。每一部分都涉及3个方面：主要内容，写作程序以及写作中常见的问题或困难。

第一节 “引言”的撰写

引言这一章由3部分组成：1）研究的总体描述；2）研究的必要性；3）论文的总体框架。

对研究的总体描述最简单的方法是：“本研究旨在探索……”，“本研究要研究的问题是……”。由于是总体描述，只需列出一些大的研究问题。研究问题中的具体细节，可在第二章文献回顾结束时提出，也可在第三章研究方法开始时提出。

关于研究的必要性，可以从两个方面着手，一个是此项研究的理论价值，另一个是此项研究的实用价值。理论价值一般在于能证明某个理论，或推翻某个理论，或修改现行理论，或创立某个急需的新理论模式，或澄清某个有争议的问题，或加深对某个现象的理解。博士论文除了有应用价值以外，必须要有理论价值。正如 Rudestam & Newton（1992）所说：“博士论文仅仅解决某个实际问题还远远不够。”应用语言学研究结果的应用价值可以体现在多方面，有的能对改进外语教学提供积极的建议，有的有助于外语词典的编撰，有的可为国家外语教育政策的制定提供实证依据。在阐述此项研究的必要性时，还可以说明激发你从事该项研究的原因。值得指出的是，这一部分不是文献回顾，没有必要列出大量文献。

描述论文的总体框架时，要包括每一章的标题。如果论文是两个层

次的模式，即："部分"是第一个层次，"章节"是第二个层次，描述时这两个层次都要包括在内。有时引言这一章还包括本项研究涉及的关键术语界定和关键术语之间的区别和假设（Rudestam & Newton，1992）。更为常见的做法是，关键术语的理论定义放在文献回顾这一章，它们的操作定义，放在研究方法这一章。

尽管引言是论文的第一章，但是应该在完成其他章节以后再撰写。原因很明显，等到写研究结果与讨论这一章时，原来开题报告中的研究问题可能已经修改过多次。

第二节　"文献回顾"的撰写

文献回顾能够体现论文作者对研究问题的理性认识、论证研究必要性的能力，以及宏观综合文献的能力。这一章是整个论文中最难写的一部分，通常比想像的要费力得多。要写出高质量的文献综述，必须要经过这一艰辛的过程，也只有这样才会对自己从事的研究产生全面深刻的理解。令人遗憾的是，许多研究生，甚至一些导师也不太注重这一部分的写作。有些研究生认为这一章是论文格式的需要，有的甚至认为增加这一章可以满足论文长度的要求。他们常常在最后一刻把所有的定义和手头文献资料随意地拼凑在一起、不管它们之间是否存在内在的逻辑关系。坦率地说，在过去的硕士论文中，有相当一部分的文献回顾这一章，质量比较差。学会如何写文献回顾是研究生学习非常重要的一部分，它需要师生共同努力。

1. 文献回顾的主要内容

从广义上说，任何研究都不应该是孤立进行的，它必定与过去和未来的研究联系在一起。文献回顾就是要为读者提供这样的背景知识。这一章不仅要描述这项研究所涉及的研究领域的历史与现状，而且要解释这项研究的原创性与承继性体现在何处。例如你需要说明，你的研究与前人研究有哪些不同之处？你是怎样在吸收别人研究设计的基础上，形成自己的研究设计？这项研究设计是以何种理论为基础？等等。读完这一章后，读者应能得出结论：你的选题值得研究，你的研究设计比较合理。

具体地说，文献回顾应包括以下几个方面：

1）关键术语的理论定义；
2）从理论的角度对选题进行审视；
3）对相关的实证研究进行评价；
4）对相关研究设计（包括相关的研究工具）给予批判性评价；
5）研究的理论框架。

以上5个方面的内容，一般包含在文献回顾中，但在具体撰写时不一定按照这一顺序，同时文献回顾的内容也可以根据各自的写作目的和个人爱好进行适当的增减。

1.1 给关键术语下理论定义

前面第二章曾经提到，研究中所有的变量都必须下理论定义和操作定义。在文献回顾这一章要给关键术语下理论定义。最简单的方法是把所有相关理论定义与参考文献一起列出，然而，这并不是最佳方法。要使定义更加令人信服，最好对不同文献的不同定义进行简要评价，并解释自己的定义是怎么产生的，或者为什么采用这个定义，而不采用其他定义。

有人写这一部分时，用很长篇幅列出十余个定义，不加任何评论，然后写上自己的理论定义。我们不提倡这种做法，原因是列出的定义与该项研究涉及的定义没有必要的逻辑联系。要帮助学生克服这个问题最常用的方法就是：让他说明这些列出的定义与他自己定义的关系。另一个常见的问题是有些被定义的术语并不是关键术语。例如：在研究使用交际策略与外语能力之间的关系时，他们往往先给“交际”和“交际能力”下定义，然后再给“交际策略”下定义。你可能要问：“哪些术语该下定义？哪些术语不需要下定义？”最简单的答案是：凡是研究中测量的变量都需要给予理论定义。

1.2 从理论的角度回顾文献

研究问题往往来源于某些理论，或与某些理论问题有关。例如：人们对语素习得的研究是受到Krashen自然顺序假设的启发。Wen对使用外语学习策略的研究是受到学习者作用新理念的启发。最为常见的文献回顾是首先解释与自己研究有关的某些理论或某一有争议的问题。

1.3 评价相关的实证研究

在文献回顾中，评价相关的实证研究必不可少，然而，这种评价并不需要涉及研究的每个细节。由于评价并不是为了判断相关研究的质

量，而是要证明自己研究的必要性，因此，评价不必面面俱到。这里至关重要的是根据评价的不同目的、对相关研究进行分类。例如：回顾相关的研究是为了说明研究这个问题的广泛性；有些是为了显示研究结果相互矛盾；有些是为了说明研究方法上的问题；有些是为了说明研究样本上的问题，等等。一旦对研究目的做到心中有数，你就能够确定评价的重点和范围。

1.4 审视研究设计

当你的研究与前人的研究在研究设计上有差异时，这一部分内容就必不可少。这种差异可能体现在研究对象的数量上，研究对象的类型上，研究工具上，数据收集的程序上、或者数据分析的程序上等。例如：以前关于这个问题的定量研究只采用了相关分析和多元回归分析的方法。而你的研究既用定量分析的方法，又用定性分析的方法，而且，在定量分析方面，你不仅采用了多元回归分析，还运用了其他的分析方法。这样，你需要回顾相关研究使用的研究设计，并指出它们的不足。

如果在研究工具上有差异，你的文献回顾必须对现行的研究工具进行简要描述，并加以评论、指出存在的问题。最后需要强调，你的研究将使用新的研究工具，以避免过去的问题。

1.5 对研究中理论框架的描述

这一部分对硕士论文可有可无，而对博士论文必不可少。理论框架实际上是在文献回顾的基础上形成的一系列假设。这也是假设产生的基础。例如：Ma 在研究外语学习者的语言变量与外语写作能力的关系时，提出了如下理论框架（Ma，1998：45）：

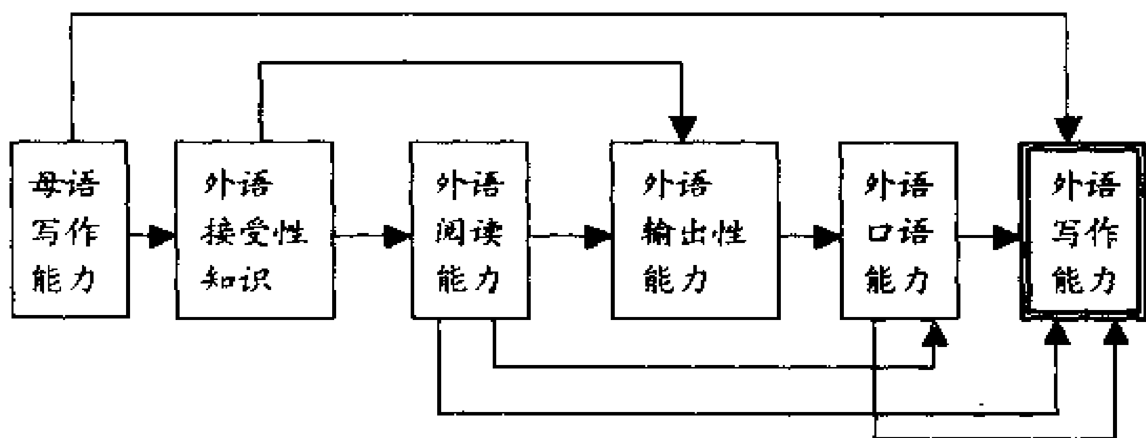


图 14.1：影响外语学习者写作能力的理论模式

Ma 在其论文中解释说："这一理论模式确定了研究的重点和 3 个因果关系，即：母语写作能力与外语写作能力的因果关系，母语写作能力与外语水平变量之间的因果关系，以及外语水平变量与外语写作能力之间的因果关系。"（p.45）

对于那些探索型开放式定性研究，理论框架只可能从实证研究的结果中产生，因此对定性研究的理论框架描述一般安排到研究结果与讨论这一章。

2. 写作步骤与常见困难

这一部分将讨论撰写文献回顾的具体步骤与常见的困难和问题。

2.1 步骤

文献回顾通常包括简短的引言和其他几个小部分。引言介绍文献回顾的范围和顺序排列。其他几个小部分都有不同的小标题，彼此之间有内在的逻辑关系。

2.1.1 列出提纲

在写文献回顾之前，应该先列一个提纲。从小标题的顺序，可以看出如何一步一步地论证。一般可以有两三个层次。提纲的主要功能是为了从宏观上把握整个文献回顾的总体方向，因此，这一阶段不需详细的细节。

一般情况下，提纲需要反复修改。如果过于匆忙，以后会发现写下的许多东西不够切题，最后还是要删除掉。即使经过反复修改的提纲，当时自己感到满意，导师也满意，但在写作过程中还可能需要修改，这一点一定要有足够的思想准备。

2.1.2 拟稿

列好提纲后，就可以进入拟稿的阶段。拟稿时，不必严格按照所列提纲的顺序来写。可以看哪一部分最熟悉，就写哪一部分，先易后难。这样一个策略可以提高写作效率，增强自信心。随着写作的深入，原来感到困难的东西，后来说不定就迎刃而解了。

2.1.3 修改

不管文献回顾的草稿写得多么好，它总不可避免地有这样或那样的漏洞。根据经验，起码要进行 3 次修改。第一次修改是宏观结构。在修改时，要反复问自己："这一部分与我的研究是不是相关？""这一部分

与其他部分有什么联系?”这一阶段的重点是整个宏观结构，不必拘泥于一些细节，如：语法错误、拼写错误等，因为如果整段有问题，将来可能要全部删掉。第二次修改的重点在段落这个层次，即段与段之间的关系。第三次修改着重在句子这个层次，看看句子与句子之间的逻辑关系和句子内部的问题。确定每次修改的重点可以大大提高修改的效率。假使导师要求学生同时改正从宏观到微观所有的错误，可以想像这其中的难度。实际上，学生在修改过程中往往重视较低层次的问题，而不重视较高层次的问题，做了很多无效劳动。因此，我们认为修改应该从宏观到微观，每次修改应有一个重点。这一规则也适合修改论文的其他部分。

2.2 常见的困难

研究生在这一阶段常会遇到各种各样的困难，其中有的困难与内容有关，有的与结构有关。

2.2.1 内容方面的困难

撰写“文献回顾”这一章时，一般需要阅读许多材料。常见的困难是，论文作者感到要回顾的东西太多，不知如何取舍。他们试图把所读的一切都放到这一章，其结果是内容杂乱，很多与研究问题无关。文献回顾不是炫耀作者背景知识的地方，而是需要论文作者根据自己研究的目的，将有关的知识，有机地组织起来，综合成一篇令人信服的说理文章。为了避免不相关、不重要的材料，作者可以把小标题按照与自己研究的相关程度一一列出来，同时确定其应有的长度。例如：Anne 打算研究中国非英语专业的学生学习动机与外语成绩之间的关系。这是一个古老的选题，已经研究了几十年。面对大量的文献，她不知所措。经过一番思考和与导师讨论，她意识到没有必要把教育领域内所有的实证研究进行详细的回顾，也没有必要回顾母语研究领域内中小学生学习动机的研究。因此她决定把文献回顾的重点放在国内外成人学习外语的动机，同时简要总结儿童和青年的外语学习动机。

跟前面的情况相反，另一种困难是，作者感到“文献回顾”可用的相关实证研究太少。当然，这不能成为免除回顾实证研究的充足理由。所谓文献的相关性，这是一个相对概念。无论现有文献的多少，我们总能找到可以评价的相关资料。否则我们怎么能证明该项研究与以往研究的联系呢？设想某论文的题目是：母语在中学生外语写作过程中的作用。可能文献中没有在外语写作过程中使用母语的实证研究，但是考察

母语在外语阅读过程中作用的实证研究很多，或者口语表达过程中类似的研究也很多。如果是这样的情况，我们可以扩大搜寻资料的视野，回顾与外语阅读和口语方面的相关研究。相关文献多的时候，就缩小搜寻视野，选择最为相关的材料；相关文献少的时候，就扩大视野，适当展开。

2.2.2 结构方面的困难

撰写文献回顾这一章对作者宏观上的综合能力要求很高。学生常常碰到的困难是，总体结构脉络不清晰，其中的小标题不能形成连贯的整体。换句话说，小标题之间缺少逻辑联系，有拼凑之感。笔者建议使用下列一般规则来组织小标题。

1）从一般到特殊；

2）从抽象到具体；

3）从理论到实证；

4）从远到近；

5）从他人的研究到自己的研究。

Cone & Forster（1993）曾用漏斗做比喻来说明文献回顾是由大到小，即文献回顾“先是总的情况，然后越来越具体，最后是集中在对现行研究的具体评论以及本项研究的基本原理”（p.105）。

结构上的另一个困难常常体现在对相关实证研究的描述和评论上。有些研究生不知道按照何种顺序组织相关的资料。下面是我们依据 Cone & Forster（1993）的看法，对克服这一困难提出的6点建议：

1）按照变量的顺序。例如：Wen（1993）在研究学习者可控因素对外语成绩的影响时，先列出学习者可控因素：动机、管理策略和外语学习策略，文献回顾时也是按照这一顺序进行。

2）按照研究设计的顺序。在定量研究中，Wen 的文献回顾顺序是由弱设计到强设计。具体地说，问卷调查先于实验研究；相关分析先于多元方差分析；共时研究先于历时研究。如果既有定量研究又有定性研究，定量研究先于定性研究。用这种方法，可以把一些不太重要的研究集中起来一起描述，然后重点评价与你研究最为密切相关的文献。

3）按照研究对象的顺序。研究对象可分为二语学习者和外语学习者。在二语学习者和外语学习者中，既可以根据学习环境把他们进一步分为在正式学习环境里的学习者和在非正式学习环境里的学习者；也可以根据年龄把他们进一步分为儿童、青年和成人学习者；还可以根据教育程度把他们进一步分为学龄前儿童、小学生、中学生和大学生。

4）按照收集数据设计的顺序。数据可以通过访谈、观察、有声思维、日记和问卷等方式收集。比方说，你的研究打算通过观察来收集数据，而前面的研究没有这样做。你之所以选择观察作为收集数据的方式，是因为访谈、日记和问卷只能获得自我报告的数据，这些数据可能在某个方面被歪曲。这样，可按照收集数据设计的顺序进行撰写。

5）按照研究结果的顺序。通过比较相关研究，找出不同研究的结果中表现出的共同点与不同点。如果以往研究的结果相互矛盾是你研究的原因，你可能要阐明这些结果相互矛盾的地方。

6）按照理论假设的顺序。例如：由于不同的人对外语水平有不同的看法，因此在相关的实证研究中对外语水平有不同的界定。如果你进行研究的原因之一是对普遍接受的观点提出质疑，用这种方法进行实证方面的文献回顾效果比较好。

很显然，选择什么方式来进行文献回顾绝不能随意决定，它要能强调文献某一方面的重要性，而强调的这一方面正是你的研究力图改进的地方。

2.3　常见的问题

文献回顾中常见的问题有两个：1）文献回顾内容与论文关系不大；2）没有适当的标题。下面我们来逐一讨论这两个问题，并提出解决这些问题的建议。

2.3.1　文献回顾内容与论文关系不大

根据我们指导研究生论文的经验，在文献回顾中常常有大量与论文无关的内容。有些人有个错误观念：把文献回顾的长度与论文质量画等号。当导师要求他们删去多余部分时，他们常常极不情愿或感到非常痛苦。判断文献回顾质量的首要标准是相关性。如果有大量不相关的内容，必须修改，删除一切无关的内容。

2.3.2　没有适当的标题

很多研究生在写这一部分时，不太喜欢使用标题，有时候可能连续数页都没有标题。由于人们的注意力和记忆力有限，如没有标题，读者很容易忘记前面读过的内容。作为作者，我们要时时为读者着想，尽量给读者提供方便。有了适当的标题，读者很容易抓住作者的思路。

当然，标题太多了也不好，因为这样可能会分散读者的注意力，并打乱他们的思路。那么，究竟多少标题为宜？这可以根据作者自己的阅读耐受力和需要来决定。

标题的数量要适当，标题的措辞也不能忽视。常有这样的情况：某一部分有 3 个小标题，第一个小标题以动词开头，第二个小标题以现在分词开头，第三个小标题又以形容词开头。很显然，同一个层次的 3 个小标题措辞不同。当然，在写作过程中，由于这些小标题分散在不同的地方，很难发现问题。一旦把这些小标题放在一起，这些问题就暴露出来，修改起来也就比较方便了。

第三节 “研究方法”的撰写

研究方法这一章要详细描述研究设计的方方面面。描述应该清晰、明了、翔实。读者看过这一章后，可详细了解整个研究过程，也可依据所描述程序，重复此项研究。与文献回顾相比，这一章相对比较容易撰写，因为一切内容都是研究者自己亲身经历过的。但是要想做到清晰、明了、翔实还是有一定难度。下面我们先来看一下研究方法这一章的主要内容，然后说明如何按照要求撰写内容的每一部分。

从表 14.1 可以看出研究方法这一章的主要部分。如前所述，研究可使用不同的设计，如：问卷调查、实验研究、个案研究或几种设计结合在一起。尽管研究设计不同，但描述每个研究设计的标题基本相同。如果一项研究使用了两个研究设计，常见的方法是把两个设计分别进行描述。换句话说，如果采用了问卷调查和个案研究两种设计，可以分别表述，先写问卷调查，再写个案研究。

下面我们将说明如何撰写研究方法这一章的每个部分，其中会特别考虑不同类型设计的差异。

引言
研究问题或假设
研究对象
研究工具
数据收集
数据分析
研究的局限性

表 14.1：“研究方法”这一章的主要内容

1. 引言

研究方法的引言简要介绍研究设计和这一章的总体框架。作为对一项研究设计的总体描述，这里只需要一两句话加以概括。例如：为了弄清学习策略与专业英语四级成绩之间的关系，本研究对来自5所大学的300名英语专业的学生进行了问卷调查。你也可以这样写，本研究使用问卷调查的方式来探索使用语言学习策略与外语成绩之间的关系。总体框架是让读者了解这一章的大致情况，其描述只限于第一层次的标题，不必提供详细内容。

2. 研究问题或假设

在这一部分，总的研究问题和具体的研究问题一起提出来。如果研究人员已有了研究假设，就把所有的假设列出来。撰写这一部分有两个目的：第一是让读者了解要研究的问题，第二是为后面“研究结果和讨论”这一章提供框架。换句话说，所有在研究方法这一部分出现的研究问题在研究结果这一章都要一一回答。然而，研究生在写研究结果这一章时往往不能和研究问题一一对应，有时报告研究结果与研究问题的顺序也不一致。

与研究方法的其他部分不同，所有的研究问题或假设都用一般现在时，而不用一般过去时。研究问题中的措辞要准确明了。

这一部分的研究问题和假设往往与论文开题报告中所提出的研究问题和假设有些出入，原因是多样的。原因之一是，研究人员在反复分析数据的过程中发现新问题。例如：你的假设可能是冒险与外语成绩没有关系，你的研究结果证明不能排除零假设。于是你选择年龄作为调节变量做进一步的研究。这样，你在原来的研究问题中就要增加一个问题：年龄会影响冒险与外语成绩之间的关系吗？与上述情况不同的是，你也可能删掉一两个最初提出的研究问题，因为你后来意识到，如果你把所有的研究问题都放在论文里，论文的内容就过于庞杂、冗长。

3. 研究对象

研究对象通常是指那些被研究个人或群体，他们的活动能为所研究的问题提供可以测量的、有价值的数据。研究对象可能被邀回答问卷或接受访谈，也可能成为被观察的对象。在实验研究中，一般用参与者

（participants）来描述研究对象和实施实验的人。“参与者”比“研究对象”的含义更为广泛，参加实验组和对照组的人属于实验的参与者，负责实施实验处理的教师或研究人员也是实验的参与者。研究对象这一部分要写明研究对象的人数、背景情况、挑选程序。记录这些情况很容易，但当研究对象的人数中途发生变化时，情况就变得有些复杂。例如：在研究使用语言学习策略与外语成绩之间的关系中，一些研究对象回答了使用语言学习策略的问卷调查，但没有参加英语水平考试，还有几个研究对象忘记写自己的学号，这样跟水平考试的分数无法对应。在历时性研究中，有些第一次参加的研究对象，由于种种原因第二次不能参加。凡是遇到研究对象有中途流失的情况，论文中都必须如实报告研究对象的变化，并说明其中的原因。

研究对象的个人情况包括性别、年龄和教育程度。如果他们的家庭背景、在何地上中学、在何地上大学等变量对因变量产生影响的话，它们也应包括在内。为了简便起见，有关研究对象的情况可以列一个表，再进行简要的文字叙述。记录选择研究对象的程序必须翔实，以方便研究人员重复操作。如果采用了随机抽样，需要清楚说明是哪一种随机抽样：是简单随机抽样？系统随机抽样？还是分层随机抽样？如果使用的是便利样本，务必解释清楚选择样本的原因。你的说明可能是：选择这个班级的原因是教那个班的老师是你同学，或是因为你本人教那个班。有时研究对象是自愿者，你也要说清楚，如果最后给志愿者小礼品以示感谢，这一点也需写明。

在数据分析中，如果根据考试成绩或根据对某些变量的回答，对研究对象重新分组，那么，关于重新分组的情况以及分组的标准都要在这一部分提供。

撰写这一部分时，研究生常犯的错误是不知道研究对象中该包括哪些人。例如：在问卷调查研究中，有人把改试卷的老师也看作研究对象，有人把设计试卷的人也看作研究对象。鉴定研究对象的最简便方法是检查他们的活动能否直接为研究问题提供有价值的数据。如果能，他们就是研究对象，否则就不是研究对象。另一个常见的问题是，这一部分写得不够详细，比如：有人没有解释选择方便样本的原因；有人没有说明研究对象人数中间发生变化的情况；还有人没有说明选择志愿者的过程。

4. 研究工具

研究工具这一部分要介绍本项目中用于测量变量的工具。例如：用语言测试来测量研究对象的外语水平，用问卷来调查研究对象对外语学习的态度，用访谈来了解研究对象学习词汇的策略等等。如果这一研究工具以前已经使用过，必须描写其内容、类型、信度和效度。如果是自己设计的研究工具，必须把设计的步骤以及小规模先导试验研究的结果详细描述。这里还应该包括自变量和因变量的操作定义。有关调查问卷这一研究工具的范例，请参见附录一。

5. 数据收集

数据收集这一部分要详细介绍数据收集的情况。这一部分要写清楚：什么时候收集数据？谁负责收集数据？如何收集数据？在哪儿收集数据？关于什么时候收集数据，不仅要写明具体日期，还要写明是课内，还是课外。关于谁负责收集数据，要说明是研究者本人，还是其他人。如果是其他人，还要说明他们是否胜任，有没有对他们进行培训。如果培训了，培训又是如何进行的？关于数据收集的方式，也要详细叙述。是亲自收集的还是通过邮寄收集的？是个人访谈还是小组访谈？访谈时有没有录音？此外，收集数据的环境是否安静？是否舒适？等等。

如果测量的自变量和因变量有多个，测量方法的描述应该逐个进行。如果是跟踪研究，要说明前后不同时间的测量是否相同。

有些研究需要对受试者或研究助手进行培训。例如：为了收集有声思维的数据，必须为受试者提供系统、有效的训练。当多个访谈人员参加同一个研究项目时，必须对访谈人员进行培训，否则访谈内容和方式的统一性就难以保证。无论是受试者，还是研究助手的培训，培训的程序和效果都必须在数据收集这一部分详细报告。

6. 数据分析

数据分析这一部分需要报告分析数据的程序及其信度。如果是定量分析，先要说明使用统计软件的名称，简要介绍统计分析的主要步骤和程序。接着需要比较详细地描述数据分析的程序。这里以问卷数据分析为例。描述要涉及以下几个问题：如何处理缺值？哪些问卷题的数值进行了转换？分类的信度指数怎样？哪些类别删除了？最后要报告统计数据，并

用表格形式把分析的类别以及题目数和信度指数统统列出来。

如果是定性数据分析，应该说明分析的程序，解释类别产生的标准和过程。此外，需要报告研究者为提高定性数据分析的信度所采取的措施。例如批改的作文分数是个因变量，这里就必须告诉读者作文的评分人是谁，两个评分人的相关系数是多少。

7. 研究的局限性

研究过程中难免会有这样那样一些应该满足但由于种种原因未能满足的条件，由于这些因素的存在导致研究存在一定的局限性。研究人员在"研究的局限性"这一部分应该如实告诉读者该项研究所存在的不足之处，提醒读者在使用该项研究结果时持谨慎态度。局限性可能体现在研究设计的不同方面。有的可能在样本中，如：样本太小、样本没有随机抽样、参加者都是志愿者等；局限性也可能在收集数据中，如：数据缺值率太高；局限性还可能在测试中，如：尽管英语专业四、八级是全国惟一的权威性的测试，但我们并不知道它们的信度和效度。

研究的局限性也可以放在最后一章结论中阐述。这样做的好处是可以详细讨论研究的局限性可能对研究结构内外部效度产生的影响。

第四节　"研究结果与讨论"的撰写

这一章最常见的写法是，一边报告研究结果一边讨论。也就是说，把研究结果与讨论结合在一起写。这样合起来写至少有两个好处：一是避免在讨论时重述研究结果。另一个好处是，读者容易理解作者的思路。

也有人把研究结果与讨论分开来写，也就是说，这一章的前半部分报告研究结果，后半部分讨论研究结果。赞成这种做法的人认为，两个独立的部分比较清楚，也比较容易操作。这两种做法都是可行的，但第一种做法更为自然，更为普遍。

1. 报告研究结果

这一节讨论 3 个问题：1）什么是研究结果？2）如何安排"研究结果"的布局？3）如何报告定量和定性的研究结果？

1.1 什么是研究结果?

所谓研究结果是指与研究问题直接相关的研究发现。有些研究发现可能与研究问题没有直接关系，因此这些发现不必写在研究结果里。初涉研究的人往往把数据分析的结果与研究结果混为一谈，他们的报告不是围绕研究问题。相反，他们逐一描述数据分析的结果，而把研究问题放在一边。为了避免上述问题，研究结果的报告自始至终不能偏离研究问题，换句话说，凡是与研究问题无关的数据分析结果都不需要在论文中出现。

1.2 如何安排“研究结果”的布局?

报告研究结果的最佳布局是根据研究问题的顺序。为了方便读者，在研究结果这一章的开头部分把研究问题重述一遍。此外，这一章的小标题最好能描述研究问题的内容，这样，通过阅读小标题，读者就可把具体的研究结果与相关的研究问题联系起来。

在这一部分的布局安排上，常见的一个问题是，有人常常根据统计分析的类型来逐一报告结果，比如他们用 t 检验、方差分析或多元回归作为报告研究结果的小标题。这些小标题从表面上看似乎清楚、明确。然而，统计分析只不过是回答研究问题的工具，并不能显示研究结果怎么与研究问题相联系。另一个常见问题是研究结果的顺序与前面提出的研究问题顺序不相同，有的甚至不仅顺序不一样，连题目也改变了，这样论文就缺乏前后的一致性。

不少研究既有定量研究的结果，又有定性研究的结果。那么，如何报告这两种研究结果呢? 首先，要确定这两种研究结果与研究问题的关系。如果定性研究的结果是用于进一步说明、补充和支持定量研究的结果，同时这两种研究结果都是针对同一研究问题，那么，定性研究的结果最好与定量研究的结果一起报告，其顺序为先报告定量研究的结果，后报告定性研究的结果。如果这两种研究结果是针对不同的研究问题，可以分开报告，也可以合起来报告。

1.3 如何报告定量研究的结果

一般说来，定量研究的结果可用图表加文字叙述的方法来报告，这两种不同形式的结合让读者既得到数字信息，又得到文字信息，两者可以互为补充。

用文字描述定量研究的结果通常有 3 种方法（Rudestam & Newton, 1992)。第一种是标志性陈述，向读者表明哪张表是有关哪个研究的结

果。第二种是专业性陈述，它是从专业的角度来表述研究结果。第三种是非专业性陈述，即用非专业的语言来说明研究结果。下面是一项假设研究的3种基本陈述。

标志性陈述

表14.2是3组学生使用不同策略的平均分和标准差。

小组名称	管理策略		母语策略	
	平均分	标准差	平均分	标准差
高分组	4.25	.464	2.32	.521
中间组	3.52	.521	3.05	.563
低分组	3.05	.400	3.68	.501

表14.2：3组学生使用管理策略和母语策略

专业性陈述（实际发现）

如表14.2所示，在3组中高分组对管理策略回答的平均分（4.25分）最高，对母语策略回答的平均分（2.32分）最低。低分组对这两个策略的回答与高分组正好相反，对管理策略回答的平均分（3.05分）最低，对母语策略回答的平均分（3.68分）最高。中间组介于两者之间，具体地说，对管理策略回答的平均分（3.52）比高分组低，比低分组高，而使用母语策略的平均分（3.05）比高分组高，比低分组低。

非专业性陈述

在3组中，高分组使用管理策略最多，低分组使用管理策略最少，中间组介于两者之间。与之形成对照，低分组使用母语策略最多，高分组使用母语策略最少，中间组介于两者之间。

如何有效地使用插图和表格来报告研究结果？有两点需要考虑：一是如果表格能说明问题，不要用图，因为图占用的空间大。二是在信息处理方面，表中的列比行更便于读者阅读归类。换句话说，如果要比较一组数字，最好是用列的形式，而不用行的形式。关于如何制表，详见第十五章的表格这一部分。

1.4　如何报告定性研究的结果

在报告定性研究的结果时，主要以文字叙述为主，偶尔辅以图表。与报告定量研究结果的3个基本陈述不同，报告定性研究的结果因研究

而异，没有统一的格式。

对于单个案研究，需要围绕研究问题详细叙述个案。假如个案研究是儿童外语发展阶段，需要对其发展的每一个阶段进行描写，并举例说明。在多个案研究中，报告的重点应是这些个案之间的相同点与不同点。同时，需要从面谈、有声思维、日记等中，大量直接引用研究对象的原话，或者间接重述研究对象说过的话或干过的事，这样才能使它们之间的相同点与不同点具体化。

尽管报告定性研究结果的方法多种多样，但它们有一点是相同的：必须进行分析与综合。因为没有任何一个读者愿意阅读一段接一段的引语，而不知道引用这些内容的目的，以及这些引语与研究问题的关系。有时候，报告定性研究的结果也可运用图表。例如：儿童外语发展的不同阶段可用一个流程图来表示。在 Wen & Guo（1998）的研究中，使用母语造句的动态功能就是用图来表示的。

2. 讨论研究结果

每报告一个研究结果，接下来都要进行适当的讨论。讨论的目的一方面是为了阐述你的研究与现有文献之间的联系，另一方面为了展现你对所研究问题的全面了解。然而，不少研究生的研究结果讨论这一部分写得不够理想，影响论文的质量。其中一个重要原因是他们觉得这一部分没有东西可写。有些人甚至分不清报告研究结果与讨论研究结果的区别。下面首先说明什么是讨论研究结果，然后对全面、恰当讨论研究结果提出建议。

2.1　什么是“讨论研究结果”?

在讨论研究结果时，你要对研究结果发表自己的观点或看法。具体地说，你要说明产生这一研究结果的可能原因、这一研究结果的意义以及对未来这一领域研究的建议等等。报告研究结果与讨论研究结果最本质的区别是：前者是陈述事实，后者是阐述观点。观点的阐述必须建筑在现有理论或前人的研究的基础之上，至少要符合一般人的常识。因此，在讨论中，引用文献必不可少。一般说来，研究结果讨论不能太长，不然的话会分散读者的注意力。如果有必要，详细的讨论可在最后一章结论中进行。

2.2　如何“讨论研究结果”?

当读到具体的研究结果时，人们自然要问的一个问题是：“为什么

有这样的研究结果?”因此，讨论研究结果首先要考虑对研究结果的解释。解释时不能凭空乱想，逻辑上一定要站得住脚，也就是说，所进行的解释必须要有现有理论和其他实证研究的结果为支撑。最缺乏说服力的解释来源于直觉。此外，解释原因时不能只是一面之词，要考虑多种可能的原因，同时还要顾及研究本身的局限性。例如 Wen（1993）在研究中发现女生在英语专业四级考试中的成绩比男生成绩好。她在讨论这一结果时写道：

“造成这样强相关的原因可能不止一个。一种解释是，女生在学语言方面比男生更有天赋，而男生在数理化方面比女生更有天赋。这一点已被许多研究所证实，如：Allen & Valette（1977），Farhady（1982），Larsen-Freeman & Long（1991），Moccoby & Jacklin（1974），Nyikos（1990）。另一种解释是，学英语专业的男生在整个男生中不是最优秀的。按照中国的传统，最好的男生通常学习理工科。如果接受这个观点，我们就要对女生在学语言方面比男生有天赋提出质疑（pp.143 – 144）。”

2.3 对研究结果意义的评价

对研究结果意义的评价也属于讨论这一章的范畴。其意义既有理论方面的，也有实践方面的。在对理论意义评价时，要把研究结果与前人的发现或现有理论联系在一起，指出这一研究结果与前人的研究结果相同，还是相左。如果相同，这就增加了前人研究结果的信度。如果不同，就需要给出各种可能的解释。如若不是研究方法导致了结果相左，那就增加了我们对这一问题认识的复杂性，同时会丰富现有的理论。

研究结果常常不仅有理论意义，而且有实践意义。在指出其实践意义时，必须明确外语教学如何得益于这一研究结果。换句话说，可根据研究结果，为改进教学提出具体建议。当然，所提建议必须来自研究结果，如果所提的建议与研究结果无关，尽管它们对教学有益，也必须删除。

2.4 对未来研究的建议

讨论研究结果的下一步是为未来研究提出建议。假如你的研究与前人的研究不一致，你所推测的原因是：1）你研究的样本不够大；2）所研究的对象存在差异，比如，你的研究对象是成人，而前人的研究对象是儿童。凡此种种，你可以建议未来的重复研究可以选择更大的样本，研究对象的范围有所扩大。同样，对未来研究的建议也必须和研究结果相关。

2.5　讨论研究结果中常出现的问题

在研究结果与讨论这一章中，那种没有适当的讨论，只用几个图表就简略报告研究结果的做法是最不可取的。要知道，讨论研究结果的过程，是充分展现研究人员研究素质的过程。如果这一部分内容欠缺，读者会认为研究人员的背景知识有限，对研究问题的理解肤浅，论文缺乏深度。另一个常见的问题是对研究结果的解释非常简单，有时甚至不合乎逻辑。为了克服这一毛病，最好的办法是跟同学或导师一起讨论研究结果，通过讨论，加深对研究结果的认识。此外，还可对照文献回顾这一章前人的研究，指出自己的研究跟他们的研究是否相同，并给予恰当的解释。

第五节　如何写“结论”

论文看到这里，读者已经跟着你的思绪，从纷繁复杂的过程中了解了你的全部研究。尽管如此，由于记忆力所限，他们依然难以记住该论文的要点。最后一部分结论的撰写就是为了帮助读者完整地回顾论文的主要内容。有人认为前面章节的内容已经非常翔实，最后再重述一遍，难免会显得多余。这是不对的。一般来说，最后的章节往往给读者留下的印象最深。而且这一部分内容的优劣对论文质量的评估也是非常重要的。在组织论文答辩的日子里，参与答辩委员会的委员们，往往要同时阅读多本论文，在这种情况下，着重审看你的论文结论部分，根据这一部分的阐述来初步判定该论文的质量，是很自然的事。

结论部分包括3方面内容：1）主要研究成果；2）本项研究的意义；3）对未来研究的建议。

1. 主要研究成果

所谓主要研究成果，就是把最重要的研究结果进行全面总结。这里没有必要把研究结果那一章节的内容一字不漏再重复一遍。这一部分有两种写法：一是先把研究问题重复一遍，这里只需要交待总的问题，而不需要说明总问题下面的具体细节。接着根据总的问题，把主要研究成果列出来。一般的做法是根据主要研究问题的顺序，逐一报告主要成果。另一

种做法是，不受原来研究问题的限制，将主要的研究结果进行归纳，总结。

无论采用哪种方法，都不需要使用行话、术语和统计数字。这里所需要的是非专业性文字叙述，叙述必须简明扼要。除了报告与研究问题相关的研究成果外，也可以报告预想不到的研究结果。

2. 研究的意义

在前一章，讨论每个研究结果时，已经阐述了它的理论意义和实践意义。然而，那是围绕具体研究结果来阐述的，因此，可能会使读者只见树不见林。在结论这一章再来阐述该研究的意义的时候，就应该综合各项研究结果，从宏观角度进行论述，给读者一个完整的图像。要登高览胜从 3 个角度阐述这项研究的意义，即：理论意义、方法论意义和实践意义。

2.1 理论意义

文献回顾这一章评述了与研究问题相关的各种不同的理论假设。在这些众多的假设之中，不免其中有些互相矛盾。那么你的研究结果是支持了还是否定了其中的某些理论？这里不是就某个具体研究结果来评述研究的理论意义，而是评述与一组研究结果相关的理论意义。如果你的研究结果与现行理论的预测不相符，你就需要解释其中的原因。

这里，我们来看一看 Wen 的研究是如何讨论其理论意义的。限于篇幅，所引部分只讨论一个理论意义。实际上，一项研究可有多个理论意义。

在第二章的 2.3.4 部分，介绍了 Stern 的外语教学中的 3 大争端。第一大争端是母语的作用，有人极力主张外语教学中使用母语作为媒介语，相反，有人则坚决反对借助母语学外语；第二大争端是外语学习究竟是一种有意识、自觉的行为，还是一种无意识或潜意识的行为；第三大争端是学习者交际时，究竟应该把主要注意力放在语言形式的准确性上，还是语言的流利程度上。这 3 大争端是传统的学习方法与非传统的学习方法分类的基础。在外语教学与研究界，一般认为这两种方法不可调和，传统的方法产生学习效率低下，非传统的方法产生的学习效率高。

本项研究的结果从学习者的角度，对 3 大争端提出了完全不同的观

点。成功的学习者并不把传统的与非传统的学习观念看成相互对立、不可调和。相反，他们认为这两者相互补充，并在学习过程中把两者有机地结合起来。只用传统方法的学生或只用非传统方法的学生，他们的成绩不相上下。然而，无论是传统的方法还是非传统的方法都不如把两者结合起来的效果好。

以上的研究结果表明，从理论上说，对传统方法的偏见应该根除，对非传统方法的热情应该降温。Krashen 及其追随者极力主张无意识的学习可以导致二语技能的熟练掌握，而有意识的学习对熟练掌握二语的作用非常有限。显然，该项研究结果对 Krashen 等人的理论提出了挑战。然而，现在还很难说这些研究成果能应用到其他学习者身上。如果未来研究的结果与此研究结果相同或相似，Krashen 及其追随者应修改他们的理论，对传统方法与非传统方法不再厚此薄彼，提倡按照学习任务的不同性质，灵活地把两种方法结合起来使用。如果未来研究发现的结果与该项研究不同，那么他们的理论必须考虑文化因素和不同学习条件对学习方法的影响（Wen，1993：249 - 250）。

2.2 方法论意义

研究方法的发展对研究的发展也是非常重要的。你的研究结果对研究方法的发展有贡献吗？如果这项研究创新性之一是由于使用了新方法，那这一研究就是在研究方法上有了一定的贡献。例如：Wen（1993）的研究在研究方法上有两点贡献。一是她使用路径分析来研究学习者的各种因素与外语学习成绩的关系。在前人的研究中，研究人员多使用多元回归来解决类似的问题。多元回归只能预测每个自变量的相对重要性，而路径分析不仅如此，而且还能够显示自变量之间的关系。二是她的研究运用了定性分析的数据来进一步说明和补充定量分析数据的不足。结果证明，定性分析的数据比定量分析的数据有时能揭示出更多有价值的信息。这一发现说明，定量研究与定性研究的有机结合比单一的定量或定性研究更加有效。

2.3 实践意义

除了理论意义和方法论意义外，应用语言学领域开展的研究，一般还有实践意义。在陈述实践意义时，不仅要叙述这些研究结果的潜在用途，而且要进一步讨论如何实施这些建议。例如：下面是 Wen（1993：250）论文中有关实践意义的一段叙述。

实践意义

本项研究的实践意义分两个方面讨论：第一、教师如何利用该项研究结果来调整学生可控因素，以促进外语学习的效率和独立学习能力的培养。第二、教师如何根据学习方法对英语成绩的影响来帮助学生克服对学习方法的偏见。……

3. 对未来研究的建议

对未来研究的建议既包括宏观建议，也包括微观建议。换句话说，建议既要涉及研究的总趋势，又要提出具体的研究建议。下面是 Wen（1993：258）论文中"对未来研究的建议"的部分摘录：

建议研究的问题分为两类：第一类是宏观的，第二类是微观的。在宏观层面上的研究问题要回答学习者可控因素与英语学习成绩的关系。在微观层面上的研究问题要测量学习者各个可控变量，或两类变量之间的关系。

1）宏观方面的研究问题

首先，同样的研究可在更多的英语专业学生中反复进行。这些研究成果能给我们提供一幅中国英语专业大学生学习英语过程的完整图画。同样的研究还可在中学生中进行，也可在非英语专业的学生中进行。研究结果可用来描绘整个中国学习者的可控因素与英语成绩之间的关系　通过进行这一系列的研究，希望能设计出一套关于学习者因素的标准问卷，并建立影响中国外语学习者成绩的各种因素的理论模式。如果有可能，还可以跨文化进行这样的研究。从跨文化环境中获得的研究结果更有助于建立全面的外语学习理论。

第二点，……

第三点，……

第四点，……

第五点，……

总之，未来关于学习者可控因素与英语学习成绩之间关系的研究可选用更大的样本，还可进行横向/纵向研究，如果有可能，还可进行跨文化研究。……

2）微观方面的研究问题

……根据该项研究的发现，词汇学习、母语策略和含混容忍度这三个方面仍旧需要进一步的研究。关于词汇学习，未来的研究可在中学生、非英语专业的学生和英语专业的学生中进行，这些研究可把重点放在成功学习者与不成功学习者的差异上。……

建议研究的问题必须与论文中的研究相关，必须是真正需要研究的问题。有些研究生写这一部分时没有经过深思熟虑，结果他们建议研究的问题要么不重要，要么与他们自己的研究没有关系。

小结

引言这一章给读者提供研究的鸟瞰图，一般包括3个方面：1）对研究的总体描述；2）研究的必要性；3）论文的总体框架。文献回顾这一章描述研究的背景，它可能包括5个方面的内容：1）给关键术语作出理性定义；2）从某个特定的理论角度审视研究问题；3）描述并评论相关的实证研究；4）如有必要，还需要介绍并评价前人的研究方法，其中包括与研究相关的研究工具；5）描述理论框架。研究方法这一章要详细描写研究设计。具体地说，这一部分应回答以下问题：1）研究问题有哪些？2）研究对象是谁？他们如何被选择？3）研究工具是什么？如何设计？4）如何收集数据和分析数据？5）该研究有哪些局限性？对以上这些问题的描写要清晰、明了、详细，以便其他研究人员重复此项研究。研究结果与讨论这一章应该按照研究问题的顺序来安排布局。只报告那些能回答研究问题的研究结果。讨论这一部分应该对研究结果进行合理的解释，评价其理论意义与实践意义，对未来的研究提出建议。结论这一章总结主要的研究结果，阐述其理论意义和实践意义，对未来的研究提出建议。与研究结果与讨论这一章不同的是，结论这一章的撰写不要依据具体的研究问题，而是紧紧抓住本项研究的大问题。

讨论题

找一本学位论文并回答下列问题：

1. 你认为“引言”部分写得怎么样？
2. “文献回顾”这部分是怎么安排布局的？你能不能提一些修改建议？
3. “研究方法”这部分写得详细吗？
4. 如何报告研究结果？所有的研究问题都回答了吗？
5. 所有的标题和小标题都合适吗？如果不合适，如何修改？
6. 讨论充分、恰当吗？如果不行，如何改进？
7. 你认为最后一章写得怎么样？你对此有何评价？

第十五章　论文的写作格式

写论文不同于写诗歌、小说和散文。在创作诗歌、小说和散文的过程中，无论是内容还是形式，都应尽可能不要与他人雷同。然而，论文有规定的写作格式。尽管论文中所报告的研究各不相同，但写作格式基本统一。研究生必须对统一的要求有所了解，以确保论文能达到规定的要求。本章首先讲述英语论文写作格式方面的基本要求，接着简要介绍美国心理学协会对论文写作格式的有关规则。

第一节　论文写作格式的基本要求

论文写作应该简洁明了。其写作与创作有两点不同：一是谋篇布局，二是语言风格。

1. 谋篇布局

正如在第十三章开头所提到的，论文包括 3 个部分：开始部分、主体部分和结尾部分。每一部分的结构与主要内容有统一的基本要求。例如：论文的主体部分必须包括“文献回顾”、“研究方法”、“研究结果和讨论”和“结论”。作为论文作者，不可以随意增加或删除其中的某一部分。

就每一章的结构而言，每个大标题下，开篇都有一段话作为引言。在引言里，需要描述这一部分的谋篇布局，以便读者在阅读之前做到心中有数。每一章的结尾部分，都有对整个章节的小结。在段落与段落之间，都有承上启下的转接。当读者读完每一部分时，会有小标题提醒读者下面要谈论的问题。论文中运用图和表格使信息表述得更为清楚、明了。

与写侦探小说不同，论文中的每部分内容都要清楚说明写作意图，不要给读者设置悬念，让读者时时发问：“作者写这一部分用意何在?”作为论文的作者，心中要时刻装着读者，尽可能地为读者提供便利。

2. 语言

论文中使用的语言应正式、简洁。要多用被动语态，少用主动语态。尽量避免使用第一人称*，尽量不用花哨的语言和情感强烈的语言，如："太棒了！太妙了！"等。

时态是论文语言需要注意的重要方面。常用的时态是现在时与过去时。时态使用不恰当可能会违背科学伦理（Day，1989）。按照科学伦理，一旦论文正式发表，它就成为大家承认的知识。因此，在引用他人已发表的研究成果时，用现在时，以示对他人的尊敬。而在描述自己的研究时，用过去时，以示谦虚。下面是论文中运用现在时与过去时的通行做法：

1）下面情况一般用现在时：

(a) 描述研究的必要性。

(b) 描述前人在杂志上发表的研究成果。

(c) 讨论现行理论和著名的原则、原理。

(d) 描述图或表中的研究结果。

(e) 讨论自己的研究结果。

(f) 阐述研究结论。

(g) 对未来研究提出建议。

2）下面情况一般用过去时：

(a) 描述自己的研究方法。

(b) 报告自己的研究结果。

(c) 描述前人研究的步骤。

第二节 美国心理学协会论文写作规则

应用语言学领域中的研究论文用英文写作时通常按照美国心理学协会所要求的格式，这个格式简称为 APA 格式。下面从以下 6 个方面介绍 APA 格式中部分相关的常用规则：1）表格；2）数字；3）拼写；4）引用；5）引证；6）APA 引用文献目录格式。

* 从最近国际刊物上发表的论文来看，这一规则似乎有所改变，用第一人称的情况越来越普遍。

1. 表格

正如上文所提到的，学术论文经常使用表格和文字叙述一起来提供信息。一个设计得好的表格可能会比文字叙述更为有效、更为经济，这是因为表格能使读者直接感知到数据所呈现的模式以及它们之间的关系，这是文字叙述所不及的。表格一般用来呈现定量研究的成果，时而也用来报告定性研究的成果。根据《APA 出版手册》规定，每个表格由5个部分组成：表号、表名、表目、表体以及表注。

1.1　是表还是图?

一个常见的问题是：一些研究生常常分不清表和图的区别，因此，在他们的论文中经常把两者混为一谈。表是由纵横排列的单元格组成，表格内容多为数字，它能准确无误地记录、储存、管理数据；图是用点、线、面、色彩等要素描绘而成，是具备数值意义的图形，其优点是形象、直观。

1.2　如何给表和图编号

论文中的所有表和图都必须编号，根据 APA 的规定，按照表和图在论文中的顺序，分别进行编号。换句话说，表和图的序号不能混编。

假如论文中共有 30 个表格，可以从表 1 编到表 30。然而，这种编法适合于一般论文，不适合硕士论文或博士论文，这是因为在硕/博士论文写作中，增减表格比较频繁。如果表格从头到尾统一编号，这就意味着只要一处有变化，所有的表格都得跟着一起重编。因此，我们建议按照章节编号。例如：第一章的表格可编成表 1.1，表 1.2，这样依次类推。第二章的表格可编成表 2.1，表 2.2，等等。如此做法，修改起来比较容易。表格的编号必须用阿拉伯数字，不允许用表 A、表 B，也不允许用表 I 或表 II。

和表格一样，图也是按照章节编号，编号的数字必须用阿拉伯数字。例如：第一章的图可编成图 1.1、图 1.2 等。第二章可编成图 2.1，图 2.2 等。

1.3　表格名称

每个表格都有一个名称。表格名称必须简明扼要，包含表格的主要信息。比如"t 检验的结果"和"多元回归的结果"作为表格名称就不合适，因为同一个研究中可能多次使用 t 检验和多元回归分析。如果同样的表格名称出现不止一次，就容易出现混乱，失去表格名称的意义。

上述两个名称最好改为"t 检验：实验组与对照组大学英语四级考试成绩的差异"和"多元回归：变量对大学英语四级成绩的预测力"。

1.4 表格标目

标目一般在表格的上方第1行和左侧第1列。如果标目中有缩略词，缩略词的意思必须清楚、明了，否则，就要加表注，解释每个缩略词的意思。

1.5 表格主体

表格主体在绝大多数情况下是数字，有时也可以是文字。在表格中安排数据时，行和列的布局安排取决于多种因素，一要符合统计软件中对数据格式的要求，二要符合常规的视觉习惯，三要考虑便于以表格形式输出打印，四要考虑到表格空间容量。

1.6 表注

表注可分为3类：总表注、具体表注和概率表注。总表注对表格总体提供一些信息。例如对表格中缩略词的解释。根据 APA 的格式，总表注用"*Note*."表示（"Note"用斜体，加上英语的句号）。具体表注是注释表格中的具体内容，在要注释的右上角标上小写字母。如果有几个项需要加具体注释，按照字母顺序从左到右标出。概率表注分别用1个星号、2个星号和3个星号表示。1个星号（*）表示 .05；2星号（* *）表示 .01；3个星号（* * *）表示 .001。

如果一个表格需要3种注释，先用总表注，再用具体表注，最后用概率表注。例如，表15.1和表15.2分别报告了定量研究结果和定性研究结果。

报告定量研究结果的表格

学习者类型	母语水平平均数[a]	标准差	外语水平平均数[b]	标准差
第一组（传统）	73.10	5.63	90.37	4.12
第二组（非传统）	73.29	5.38	91.08	3.51
第三组（高平衡）	74.02	5.51	91.62	3.40
第四组（低平衡）	73.10	5.04	91.00	3.74

表15.1：方差分析：4组学习者在母语水平和外语水平没有差异（Wen，1993：185）

注：传统 = 传统学习者；非传统 = 非传统学习者；
高平衡 = 高平衡学习者；低平衡 = 低平衡学习者。} 总表注

a：母语水平由高考语文分数来体现
b：外语水平由高考英语分数来体现 } 具体表注

报告定性研究结果的表格

什么	为什么	怎样	什么时候
听美国之音的《杂志博览》节目	1）练听力 2）增加背景知识 3）练写作	1）记笔记 2）用自己的话把记得的东西写下来	一周三次
记忆每周碰到的生词	扩大词汇量	使用卡片	一周一次
快速阅读（简易读物）	提高阅读速度	尽可能读得快	每晚睡觉前

表 15.2：优秀学习者制定的个人学习计划（Wen，1993：194）

2. 数字

论文中的数字怎么处理？APA 格式数字表达的方式有一系列规则。

2.1 用数字还是用文字？

英语论文中的数目是用数字表示还是用文字表示？一般规则是如果数目小于 10，用文字表示，如 three，one，four 等等。如果数目大于 10，用数字表示，例如 11，13，20 等等。当然，也有一些例外。

当小于 10 的数与 10 或大于 10 的数一起连用时，这个数用数字表示，例如“5 out of the 36 second-year students（36 个二年级的学生中有 5 个）”和“They are ranked the 3rd and 15th respectively（他们分别为第 3 名和第 15 名）是对的。在上述例子中，5 没有写成“five”和 3rd 没有写成“third”就是因为他们与超过 10 的数字连在一起使用。

当小于 10 的数的绝对值用来代表时间、年龄、考试分数和等级时，一般用数字。例如“They finished the test within 2 hours”（他们在 2 小时之内完成了测试）和“Her score on the composition is 5”（她作文考试得了 5 分）。

2.2 小数

如果一个数的绝对值不可能大于 1，是一个纯小数时，如：相关系数和概率值，就不需要在小数点前面加 0，一般写成 .40 或 .05。反之，如果一个数的绝对值有可能大于 1，这时，小数点前面的 0 就不能少，例如：0.40 和 0.05。对于表格中的小数而言，在合理保留有效数字的前提下，需要进行必要的四舍五人，删除多余的小数位。

2.3 数字的复数

数字的复数可在数字后直接加 s。例如："在 20 世纪 90 年代"，用英语可表达为 "in the 1990s"，而不必表达为 "in the 1990' s"。同样，"八九十年代之间"，用英语表达为 "between the 80s and 90s"，而不必表达为 "between the 80' s and 90' s"。

3. 拼写

关于拼写，《APA 出版手册》对拼写方式、连字符和大写都做了具体说明。

3.1 美国的拼写方式还是英国的拼写方式?

根据 APA 的规定，论文的拼写必须按照《韦氏新大学词典》（*Webster' s New Collegiate Dictionary*）所确定的标准。很显然，这是美国的拼写方式。如果你喜欢用英国的拼写方式，也完全可以。惟一不方便的是文字处理软件通常默认的是美国英语的拼写方式。无论你喜爱什么拼写方式，一旦你选定，就必须前后一致。

3.2 连字符

拼写中的一个困难是复合词中连字符的使用。大部分情况，可以在词典中找到解决方法。然而，相当一部分复合词在词典上查不到。如果碰到这种情况，可以按照《APA 出版手册》提供的一般规则。第一个规则是：如果复合形容词的意思比较清楚，不可能引起误解，就不要使用连字符，例如：modifiable learner variables（学习者可控因素）和 L2 linguistic proficiency（第二语言语言能力）。第二个规则是：如果修饰名词的新词是自己新造出来的，最好使用连字符，以避免误解，例如：如果想用 different word lists 来表达"不同词汇的列表"，最好在 different 和 word 之间加个连字符，即 different-word lists，否则，容易被误解为"不

同的词汇表"。第三个规则是：如果复合形容词出现在名词前，一般要用连字符，例如：learner-centered classroom（学习者为中心的课堂教学），t-test results（t 检验的研究结果）（APA，1992：57）。相反，如果复合形容词出现在名词后，一般不用连字符，例如 The classroom is learner centered（课堂教学是以学习者为中心），results yielded by t tests（结果来自于 t 检验）。第四个规则是：如果两个或更多的词共同修饰同一个词，这时的连字符最好保留（APA，1992：57），例如 long-and short-term training（长期和短期训练），five- and ten-year plan（5 年和 10 年计划）。

3.3 大写

众所周知，一个完整的英语句子应以大写字母开头。然而，在写论文时，可能会碰到一些情况，使人不知所措。下面介绍《APA 出版手册》中关于大写的一些规则。

1) 当一个完整的句子放在冒号后，第一个词的第一个字母要大写。例如：

   We can draw an inference from the findings: The Traditional and Non-traditional approaches are equally effective but they are less effective than a combination of these two approaches.

2) 在论文和著作的标题中，实词以及含有 4 个字母或 4 个字母以上的虚词，如连词，需要大写。少于 4 个字母的介词和冠词不需要大写。如果标题中包含带有连字符的复合词，两个单词都要大写。但这一规则不适用于引用文献。

3) 当名词后跟一个数字或一个字母，表示某一顺序中的确切位置，这个名词的第一个字母要大写。例如：on Day 1 of Semester 1（第一学期的第一天），Group A and Group B（甲组和乙组），in Table 2 and Figure 7（在表 2 和图 7 中）。

4. 引用

只有当引用比自己的话更加有力、更为有效时，才去引用。引用有正面引用，即从正面来证明某个论点；引用也有反面引用，即从反面证明某个论点。如果引用少于 40 个词，不必另起一段。如果超过 40 个词，那就称为成段引用（见例 3）。大段引用需要另起一段，每一行从左边缩进 5 个空格，下面每一行跟上面的对齐，大段引用不需要引号。

不管是把引用放在段落中，还是另起一段，都要准确地把作者、日期和页码标出来（见例 2）。如果想要对引文作修改，一定要按照《APA 出版手册》的规定。

根据《APA 出版手册》的规定，可以对引文作某些方面的修改而不作任何解释。例如，改变引文第一个单词的第一个字母，改变引文句后的标点符号以避免句法错误。至于其他改动，需要对改动的原因进行解释。如果要省略引文，在句中用 3 个小点的省略号（见例 3），在两个句子之间用 4 个小点的省略号。如果要在引文添加一些内容，应该使用方括号，而不要使用圆括号（见例 3）。如果想强调引文的某一部分，可把它变成斜体，然后再在下面加下划线，紧接着后面加上一个括号，括号内写明：italics added（斜体是作者加的）（见例 3）。

例 1

> Conceptually, Gardner (1985) sees motivation as "the combination of effort plus desire to achieve the goal of learning the language plus favourable attitudes toward learning the language" (p. 10).

例 2

> The following section will discuss empirical studies on learning strategies in terms of their findings and methodological problems. If we "review the whole of the learner-strategies research, we have to say that the area is at an embryonic stage. Conflicting results and methodologies proliferate. There are few hard findings." (Skehan, 1989, p. 98)

例 3：成段引用

Rubin (1987) more explicitly states the importance of studying students' beliefs in her review of the research on learner strategies:

> ...to better understand how learner strategies come to be used, it is essential that we account for a learner's knowledge about language and his/her beliefs about the language learning process [that is, his/her views about how learning can be successful.] because his knowledge can *form the basis for selecting and activating one strategy over another* (italics added) (p. 19)

5. 引证

如果用某个作者具体的话或论点来进行引证，要注明作者的姓和出版年份。这样可从参考书目中帮助读者确定信息来源。《APA 出版手册》

描述了在不同的场合下引证应该注意的事项。下面将介绍如何引证一个作者、两个作者或多个作者合著的著作，如何引证多个作者的两部或多部著作。

5.1 引证一部著作

引证中最简单的是引证一个作者的一部著作，需要把作者的姓和出版年份放在括号中，姓和年份用逗号分开（见例1）。如果作者的姓已经出现在文本中，只需要在姓的后面用括号注明出版年份（见例2）。

例 1

It has been argued that teachers' role is to provide the students with optimal conditions which can facilitate learning so that students can achieve similar successful results (Bloom, 1976).

例 2

Gagne (1977) also noticed that adult learners were less affected by external instruction events.

如果所引著作是两个作者所著，要把两个作者都写下来（见例3）。如果所引是三五个作者合著，第一次要把所有的作者写下来（见例4），后面再引用时只要把第一个作者的姓，加上"et al."（等）（见例5）。

例 3

The disadvantage of the multiple regression analysis is that it cannot show the complex interrelations between independent variables (Bryman & Cramer, 1990).

例 4

Studies of the good language learner (for example, Naiman, Frohlich, Stern, and Todesco, 1978) have tried to identify the strategies which successful learners use (first citation) (Ellis, 1994: 37).

例 5

Naiman et al. (1978) found a similar relationship, although in this case 'effort' on the part of the learners was also associated with instrumental motivation (subsequent citations) (Ellis, 1994: 512).

当一部著作有6个或6个以上的作者，只需要写第一个作者的姓加上"et al."。如果一部著作是两个或更多的同姓作者所著，必须把作者名字的起首字母放在括号中。

5.2 在同一括号中引证两部或多部著作

如果同一作者的两部或多部著作被引证，只需要写一次作者的姓，

然后写上出版时间，其顺序是从过去到现在。例如：

Empirical studies on student learning carried out since the 70's have found that students' learning outcomes to a great extent depend on their choice of learning approaches (Biggs, 1979, 1987).

The major controversies on L2 learning, as Stern (1973, 1983, 1992) says, center around three key issues.

如果在括号中引证的是不同作者的两部或多部著作，排列顺序要根据作者姓的字母。例如：

One major finding from the earlier studies (Barley, 1969, 1970; Clement et al., 1978; Gardner & Lambert, 1972; Gliksman, 1976; Spolsky, 1969) was that learners with an integrative motivation tended to obtain better achievement than those with an instrumental motivation.

6. APA 引用文献目录格式

引用文献排序的基本规则是按字母顺序。每条引用文献的内容包括作者、出版年份、文献的题目和出版地，各部分之间用逗号隔开，以句号结束。下面谈谈写引用文献时可能会碰到的困难。

6.1 引用期刊

如果引用的是期刊，有 4 点需要注意：

1) 论文标题和副标题的第一个字母要大写。不要给论文标题加下划线，也不要在论文题目上用双引号。
2) 期刊名称主要词的第一个字母要大写，然后在下面加下划线。
3) 使用阿拉伯数字表示期刊的卷数。如果有期数，把它放在卷数后面的括号里。
4) 使用阿拉伯数字表示论文的页码，把它放在卷数后面，如果有期数，就放在期数的后面。

请看下面几个例子：

Block, E.L. (1986). The comprehension strategies of second language readers. TESOL Quarterly, 20 (3), 463-494.

Boyle, R.P. (1979). Path analysis and ordinal data. American Journal of Sociology, 75, 461-480.

Bracht, R.M. & Glass, G. V. (1968). The external validity of experiments. American Educational Research Journal, 5 (4), 437-474.

6.2 引用著作

尽管引用著作与引用期刊有些相同的地方，但也有一些不同的地

方。这里我们只说明不同的地方。

1）在引用编著时，主编的名字放在作者名字的地方，并把“Ed”放在其后的括号中，如果主编是两个或两个以上，那就用“Eds”。

2）编著书名的第一个字母要大写，书名要加下划线。

3）版数（如第二版）放在书名后的括号中；

4）先写出版地，再写出版社名。

我们来看几个例子：

Oxford, R.L. (1990). Language learning strategies: What every teacher should know. New York: Newbury House Publisher.

Marton, F., Hounsell, D. & Entwistle, N. (Eds). (1984). The experience of learning. Edinburgh: Scottish Academic Press.

Brown, H. D. (1987). Principles of language learning and teaching (2nd ed). New Jersey: Prentice-Hall, Inc.

6.3 引用编著书目中的一篇论文或其中一个章节

有时，引用文献不是某个期刊上的一篇论文，也不是整本书，而是编著或论文集中的一篇论文或其中一个章节。例如：Wen在其研究中曾引用了Ringbom1985年主编的《外语学习与双语现象》中Palmberg写的一篇“讲瑞典语的小学生在小学开英语课前知道多少英语单词?”的论文。这个引用文献应这样写：

Palmberg, R. (1985). How much English vocabulary do Swedish-speaking primary school pupils know before starting to learn English at school. In H. Ringbom (Ed.). Foreign language learning and bilingualism (pp.89-97). Abo: Research Institute of the Abo Academic Foundation.

引用这类文献时要注意两点：

1）当主编的名字不在条目的开头位置时，姓和名不要倒置。

2）所引页码放在书名后的括号里。

我们再看几个例子。

Schmeck, R. R. (1983). Learning styles of college students. In R. R. F. Dillon & R. Schmeck (Eds.). Individual differences in cognition (Vol. 1, pp. 233-279). New York: Academic Press.

Swain, M. (1985). Communicative competence: some roles of compre-

hensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.). Input and second language acquisition (pp.235-253). Rowley, M.A.: Newbury House.

6.4 从国际权威性学术期刊看引用文献格式

仔细观察一下国际3家权威学术期刊：*TESOL Quarterly*、*Studies in Second Language Acquisition* 和 *Applied Linguistics*，我们就会发现前面两个期刊基本采用的是APA格式，但也有两个不同点。第一个不同点是：如果引用文献是期刊，期刊的名称和卷数不像《APA出版手册》所规定的那样在下面加下划线，而是用斜体。第二个不同点是：如果引用文献是著作，书名也是用斜体，而不是在下面加下划线。然而，*Applied Linguistics* 的引用文献格式与APA格式差别比较大。除了上面提到两个不同点以外，还有以下几个不同点：

1）除了第一作者以外，姓不要放在名字的首字母前面，此外，第一作者和其他作者之间用"and"，而不用"&"。

2）出版年份不要放在括号里。

3）期刊论文的题目放在单引号里面。

4）在卷数和期刊名称之间没有逗号，在卷数和页码之间用分号。

5）如果引用的是编著，把"editor (s)"这个词缩写成"ed (s)"并放在括号里。第一个字母不要大写，括号后用冒号而不用逗号。

下面是这3个学术期刊的一些实例：

(1) *TESOL Quarterly*

Beck, I., & McKeown, M. (1985). Teaching vocabulary: Making the instruction fit the goal. *Educational Perspectives*, 23, 11-15.

Dubin, F., & Olshtain, E. (1993). Predicting word meanings from contextual clues: Evidence from L1 readers. In T,

Huckin, M. Haynes, & J. Coady (Eds.). *Second language reading and vocabulary learning* (pp. 181-202). Norwood, NJ: Ablex.

[选自1997年春 *TESOL Quarterly* 第31卷第一期，第140－141页。]

(2) *Studies in Second Language Acquisition*

Berg, T. (1986). The problem of language control: Editing, monitoring, and feedback. *Psychological Research*, 48, 133-144.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J.C. Marshall, & C. W. Wallesch (Eds.). *Linguistic disorders and pathologies* (pp.1-15). Berlin: de Gruyter.

[选自 2000 年 *Studies in Second Language Acquisition* 第 22 卷，第 163 页]

(3) *Applied Linguistics*

Bierwisch, M. and R. Schreuder. 1992. 'From concepts to lexical items.' *Cognition* 42: 23-60.

Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.

Kroll, J. F. and J. Curley. 1988. 'Lexical memory in novice bilinguals: The role of concepts in retrieving second language words' in M. Gruneberg, P. Morris, and R. Sykes (eds.): *Practical Aspects of Memory*, Vol. 2. London: John Wiley and Sons.

[选自 1999 年 *Applied Linguistics* 第 21 卷第一期，第 47 – 77 页]

很显然，不同学术期刊有各自撰写引用文献的格式。因此，投稿时一定要按照具体期刊的要求，否则，你的论文很可能不会被录用。关于硕士或博士论文写作，我们建议统一按照 APA 格式，切忌将几种格式混用。

小结

论文写作与创作不同，应该清晰明了，遵循一定的结构标准，使用正式简洁的语言。应用语言学领域的论文最好按照《APA 出版手册》。这个手册包含了许多常用规则，其中有使用时态、图表、数字、大写和拼写的规则，还有有关引证、引用和引用文献格式的规则。论文格式必须体现专业水准，从头到尾按照统一要求撰写。

讨论题

找一本硕士或博士论文，和你同学一起讨论，并回答下列问题：

1. 这篇论文有无标准结构？如果没有，问题出在哪儿？
2. 你认为这篇论文的语言怎么样？是不是正式？如果不正式，请举例说明。
3. 时态使用方面有问题吗？请把这些问题列出来讨论。
4. 你认为这篇论文中的表格符合要求吗？
5. 这篇论文的连字符有问题吗？如果有，能否提一些建议？
6. 你认为这篇论文的引证怎么样？如果有错误，请把这些错误纠正过来。
7. 引用方式是否得体？如果不得体，请进行修改。
8. 引用文献的格式有问题吗？能不能根据《APA 出版手册》提出一些修改建议？

参考文献

Allen, E. D. & Valette, R. M. (1977). *Classroom techniques: Foreign languages and English as a second language*. New York: Harcourt Brace.

Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 11, 256-274.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bailey, K. M. (1983). Competitiveness and anxiety in adult second language learning: Looking at and through the diary studies. In H. Seliger and M. Long (Eds.), *Classroom oriented research in second language acquisition*. Rowley, Mass.: Newbury House.

Bailey, K. M. (1990). The use of diary studies in teacher education programs. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education*. Mass: Newbury House.

Bejarano, Y. (1987). A cooperative small-group methodology in the language classroom. *TESOL Quarterly*, 21, 483-504.

Bernard, H. R. (1994). *Research methods in anthropology: Qualitative and quantitative approaches* (2nd ed.). California: Sage.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning*, 28 (1), 69-83.

Bjorklund, D. F., Coyle, T. R. & Gaultney, J. F. (1992). Developmental differences in acquisition and maintenance of an organizational strategy: Evidence for the utilization deficiency hypothesis. *Journal of Experimental Child Psychology*, 54, 434-438.

Black, K. (1992). *Business statistics: An introductory course*. New York: West Publishing Company.

Blaxter, L., Hughes, C. & Tight, M. (1996). *How to research*. Buckingham: Open University Press.

Block, E. L. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20 (3), 463-494.

Boyle, R. P. (1970). Path analysis and ordinal data. *American Journal of Sociology*, 75, 461-480.

Bracht, R. M. & Glass, G. V. (1968). The external validity of experiments. *American Educational Research Journal*, 5 (4), 437-474.

Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.

Chen, H. (2000). A study of vocabulary learning strategies used by Chinese children. Nanjing University: Unpublished MA thesis.

Chen, J. (1996). A study of discomfort in EFL classroom. Nanjing University: Unpublished BA thesis.

Cohen, A. (1987). Using verbal reports in research on language learning. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 82-95). Clevedon, England: Multilingual Matters.

Cohen, L. & Manion, L. (1991). *Research methods in education* (3rd ed.). London: Routledge.

Collins Cobuild English Language Dictionary. (1987). England: HarperCollins Publishers.

Cone, J. D. & Foster, S. L. (1993). *Dissertations and theses from start to finish* (2nd ed.). Washington, DC: American Psychological Association.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics*, 5 (4), 1961-1970.

Day, R. A. (1989). *How to write and publish a scientific paper*. Cambridge: Cambridge University Press.

Dong, Y. (1999). ESL Learning on the Internet? A case study on a joint online course. Nanjing University: Unpublished MA thesis.

Ellis, Nick. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, Rod. (1997). *SLA research and language teaching*. Oxford: Oxford

University Press.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly*, 16, 43-59.

Flower, L., Stein, V. E., Ackerman, J. M., Kantz, M. H., McCormick, K. M. & Peck, N. D. (1990). *Reading to write: Exploring a cognitive and social process*. New York: Oxford University Press.

Gardner, R. C. & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Rowley, Massachusetts: Newbury House Publishers.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitude and motivation*. London: Edward Arnold.

Glaser, B. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.

Gong, X. (1998). A contrastive study of English and Chinese compliment responses. PLA Foreign Langnages University: Unpublished MA thesis.

Goodman, Y. M. & Marek, A. (1989). Retrospective miscue analysis: History, procedures and prospects. Occasional Paper No. 19. Tucson: Arizona Center for Research and Development, Program in Language and Literacy.

Granott, N. & Parziale, J. (2002). Microdevelopment: Transition Processes in Development and Learning. Cambridge: Cambridge University Press.

Gu, Y. Q. (1997). A study of vocabulary learning strategies by Chinese EFL learners. Hong Kong University: Unpublished Ph. D. thesis.

Guo, C. J. (1997). A process-oriented cognitive account of L1 influence on L2 writing. Nanjing University: Unpublished MA thesis.

Hatch, E. & Farhady, H. (1982). *Research design and statistics for applied linguistics*. Rowley: Newbury House Publishers, Inc.

Hosenfeld, C. (1984). Case studies of ninth grade readers. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 231-249). London: Longman Group U. K.

Hu, Y. Z. (1999). *Error treatment in EFL classroom*. Nanjing University: Unpublished MA thesis.

Huang, X. H. (1984). An investigation of learning strategies in oral communication that Chinese EFL learners in China employ. Chinese University of Hong Kong: Unpublished M. A. thesis.

Hudelson, S. (1989). A tale of two children: Individual differences in ESL children's writing. In D. M. Johnson & D. H. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 84-99). White Plains, NY: Longman.

Hymes, D. (1972). *On communicative competence*. Philadelphia, P. A.: University of Pennsylvania Press.

Johnson, D. M. (1992). *Approaches to research in second language learning*. London: Longman.

Keeves, J. P. & Sowden, S. (1992). Analyzing qualitative data. In J. P. Keeves (Ed.). *Methodology and measurement in international educational surveys*. The International Association for Evaluation of Educational Achievement (IEA).

Krashen, S. (1985). *The input hypothesis*. Hong Kong: Longman Group Limited.

Kwo, O. W. Y. (1999). Reflective classroom practice: Case studies of Hong Kong student teachers. In A. Y. Chen & Maanen, J. V. (Eds.). *The reflective spin: Case studies of teachers in higher education transforming action*. Singapore: World Scientific.

Labovitz, S. (1971). The assignment of numbers to rank order categories. *American Sociological Review*, 35, 515-524.

Lado, R. (1961). *Language testing*. New York: McGraw-Hill.

Larsen-Freeman, D. & Long, M. (1991). *An introduction to second language acquisition research*. New York: Longman.

Lauer, J. M. & Asher, J. W. (1988). *Composition research: Empirical designs*. Oxford: Oxford University Press.

Liu, Z. C. (1999). Learners' attitude towards errors and error correction. Nanjing University: Unpublished MA thesis.

Longman Dictionary of English Language and Culture. (1992). England: Longman Group UK Limited.

Lu, L. P. (1997). L2 proficiency and comprehension strategy use. Nanjing University: Unpublished MA thesis.

Ma, G. H. (1998). The relationship of L2 learners' linguistic variables to L2 writing ability of tertiary-level Non-English Majors in China. Nanjing

University: Unpublished Ph. D. thesis.

Maccoby, E. E. & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford: Stanford University Press.

Malhotra, N. K. (1993). *Marketing research: An applied orientation*. New Jersey: Prentice-Hall, Inc.

Markee, N. P. (1994). Toward an ethnomethodological respecification of second-language acquisition studies. In E. E. Tarone et al (Eds.). *Research methodology in second-language acquisition*. New Jersey: Lawrence Erlbaum Associates Publishers.

Marton, F. & Saljo, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46, 4-11.

Marton, F. (1981). Phenomenography-describing conceptions of the world around us. *Instructional Science*, 10, 177-200.

Maxim, P. S. (1999). *Quantitative research methods in the social sciences*. Oxford: Oxford University Press.

Miles, M. & Huberman, A. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Miller, P. H. & Coyle, T. R. (1999). Developmental change: Lessons from microgenesis. In E. K, Scholnick, K., Nelson, S. A. Gelman & P. H. Miller. *Conceptual development: Piaget's legacy*. London: Lawrence Erlbaum Associates Publishers.

Miller, P. H., Aloise-Young, P. (1996). Preschoolers' strategic bebaviors and performance on a same-different task. *Journal of Experimental Child Psychology*, 60, 284-303.

Norusis, M. J. (1993). *SPSS for windows: Base system user's guide*. Chicago: SPSS Inc.

Nunan, D. (1994). *Research methods in language learning* (3rd ed.). Cambridge: Cambridge University Press.

Nyikos. M. (1990). Sex-related differences in adult language learning: Socialization and memory factors. *The Modern Language Journal*, 74, 273-287.

Oxford, R. & Crookall, D. (1989). Research on language learning strategies: Methods, findings and instructional issues. *The Modern Language Journal*, 73 (4), 404-419.

Paribakht, T. S. & Wesche, Marjorie. (1997). Vocabulary enhancement activities and reading for meaning in the second language vocabulary acquisition. In Coady and Huckin (eds.). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.

Patton, M. Q. (1980). *Qualitative evaluation methods*. Beverly Hills, CA: Sage.

Pica, T., Young, R. & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, 21 (4), 737-758.

Publication Manual of the American Psychological Association. (1992).

Punch, K. F. (1998). *Introduction to social research: Quantitative & qualitative approaches*. London: Sage.

Qin, X. Q. (1998). An investigation into the internal structure of EFL motivation at the tertiary level in China. Nanjing University: Unpublished Ph. D. thesis.

Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19 (2), 229-258.

Rivers, W. M. (1983). *Communicating naturally in a second language: Theory and practice in language teaching*. New York: Cambridge University Press.

Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.

Rudestam, K. E. & Newton, R. R. (1992). *Surviving your dissertation: A comprehensive guide to content and process*. California: Sage.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32, 102-119.

Schmidt, R. & Frota, S. (1985). Developing basic conversational ability in a second language: a case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn*. Rowley, Mass.: Newbury House.

Scholnick, E. K., Nelson, K., Gelman, S. A. & Miller, P. H. (1999). *Conceptual development: Piaget's legacy*. London: Lawrence Erlbaum Associates publishers.

Seliger, H. W. & Shohamy, E. (1990). *Second language research methods*. Oxford: Oxford University Press.

Shehadeh, A. (2002). Comprehensible output, from occurrence to acquisition: An agenda for acquisitional research. *Language Learning*, 52, 597-647.

Siegler, R. S. & Stern, E. (1998). A microgenetic analysis of conscious and unconscious strategy discoveries. *Journal of Experimental Psychology: General*, 127, 377-397.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology*, 25, 225-273.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale, *Microdevelopment: Transition processes in development and learning* (pp. 31-58). Cambridge: Cambridge University Press.

Siegler, R. S. & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology*, 81, 446-457.

Siegler, R. S. & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, 46, 606-620.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Smith. M. J. (1998). *Social science in question*. London: Sage.

Stake, R. (1988). Case study methods in educational research: Seeking sweet water. In R. M. Jaeger (Ed.), *Complementary methods for research in education*. Washington: American Educational Research Association.

Stern, H. H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review*, 31, 304-318.

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford University Press.

Stern, H. H. (1992). *Issues and options in language teaching*. Oxford: Oxford University Press.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded*

theory procedures and techniques. Newbury Park, CA: Sage.

Strauss, A. & Corbin, J. (1998). *Basics of qualitative research*. Newbury Park, CA: Sage.

Strauss, A. (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Su, X. J. (1996). A study of the changes of tertiary English majors' beliefs. Nanjing University: Unpublished MA thesis.

Tashakkori, A. & Teddlie, C. (1998). *Mixed methodology*. London: Sage.

Theodorson, G. A. & Theodorson, A. G. (1969). *A modern dictionary of sociology*. New York: Thomas Y. Crowell.

Vygotsky, L. S. (1986). Thought and language. Massachusetts: The MIT Press.

Wang, L. (1999). Risk-taking and English learning: A study of the risk-taking beliefs and behaviors of English majors in China. Nanjing University: Unpublished MA thesis.

Wang, W. Y. (2000). L1 thinking and L2 writing: A study of the process of L2 by Chinese university students. Nanjing University: Unpublished Ph. D. thesis.

Wen, Q. F. (1993). Advanced level English language learning in China: The relationship of modifiable learner variables to learning outcomes. Hong Kong University: Unpublished Ph. D. thesis.

Wen, Q. F. (2001). *Applied linguistics: Research methods and thesis writing*. Beijing: Foreign Language Teaching and Research Press.

Wen, Q. F. & Johnson, Robert Keith. (1997). L2 learner variables and English achievements. *Applied Linguistics*, 18 (1), 27-48.

Wu, C. X. (1998). A study of second-year English Majors' pragmatic competence. Nanjing University: Unpublished MA thesis.

Wu, J. (1998). A study of university students' argumentative writing in English: rhetorical knowledge and discourse pattern. Nanjing University: Unpublished MA thesis.

Xu, H. M. (2001). A study of the use of meta-discourse markers by Chinese EFL learners in L2 writing: A developmental perspective. Nanjing University: Unpublished Ph. D. thesis.

Xu, Y. C. (1998). Cohesion in EFL writing. Nanjing University: Unpublished MA thesis.

Ye, X. L. (1994). The role of schema and language proficiency in EFL reading. Nanjing University: Unpublished MA thesis.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, 17 (1), 165-187.

Zamel, V. (1987). Recent research on writing pedagogy. *TESOL Quarterly*, 21 (4), 697-715.

Zhu, J. (2003). Effects of input frequency on L2 learners' reading comprehension improvement and vocabulary learning. Nanjing University: Unpublished BA thesis.

Zhu, L. Z. (1998). Influence of different tasks, lengths of planning time and lengths of learning on second language oral performance. Nanjing University: Unpublished MA thesis.

文秋芳，2000，《外语口语测试与教学》，上海外语教育出版社。

文秋芳，2001，英语学习者动机、观念、策略的变化规律与特点，《外语教学与研究》第2期。

文秋芳，2003，频率作用与二语习得研究，《外语教学与研究》第2期。

文秋芳、郭纯洁，1998，母语思维与外语写作能力的关系，《现代外语》第4期。

文秋芳、王海啸，1996，学习者因素与大学英语四级考试成绩的关系，《外语教学与研究》第4期。

文秋芳、吴学顺，1998，评价语言实验室内的口语教学，《外语界》第1期。

文秋芳、吴彩霞、Lydia So，1999，对全国英语专业四级口语水平的评估，《外语教学与研究》第1期。

附　录

附录一：A SURVEY STUDY

(Note: The following is a longitudinal survey in which the same group of students were asked to respond to the same questionnaire three times in an attempt to detect developmental patterns in motivation, beliefs and strategies.)

Developmental Patterns in Motivation, Beliefs and Strategies of English Learners in China

Qiufang Wen　Nanjing University

Abstract:

The paper reports a study on developmental patterns in the modifiable learner variables (i.e. motivation, beliefs and strategies) and in their relations based on the longitudinal questionnaire data. The subjects involved in the study were 72 English majors who were admitted into Nanjing University in 1996.

Key words:

English language learners, modifiable variables, motivation, beliefs, strategies

英语学习者动机、观念、策略的变化规律与特点

南京大学　文秋芳

提要：本文运用定量研究的方法，3次跟踪调查了南京大学1996年入学的72名英语专业学生可控因素（动机、观念和策略）及其关系的变化情况。

关键词： 英语学习者、可控因素变化、动机、观念、策略

纵观动机、观念和策略的研究历史，尽管研究的广度和深度在不断发展，但几乎所有的研究都是静态研究（例如 Cohen，1998；Ellis，1994；Gardner & Lambert，1972；Huang & Van Naerssen，1985；Oxford，1990；Wen，1993；王文宇，1998；文秋芳，1995；文秋芳和王海啸，1996a，1996b），从动态角度研究的则寥寥无几。笔者曾于 1996 年就 3 个语言学习策略的变化进行过尝试性的研究，但研究没有涉及其他的学习者可控因素，更没有涉及各可控因素关系之间的变化。与以前的研究不同，本文所要报告的九五社科项目"英语学习者动机、观念、策略的变化规律"是从动态的角度，考察几乎所有学习者可控因素的变化情况*。具体要回答的问题包括：

1. 通过大学两年的学习，大学英语专业的学生在可控因素（动机、观念和策略）上会发生哪些变化？变化的特点和趋势是什么？
2. 通过大学两年的学习，各类可控因素之间的关系是否发生了变化？如有变化，变化的特点和趋势是什么？

一、可控因素的定义及关系

1. 动机、观念和策略的定义

可控因素分为动机、观念和策略（Wen，1993；文秋芳和王海啸，1996）。动机分为表层动机和深层动机（Biggs，1979）。表层动机通常与个人的前途直接相关，动力来自外部；深层动机一般不与学习者个人的前途和经济利益发生直接的联系，学习动力来自对英语语言或文化本身的兴趣。

观念指学生在学习英语过程中通过自身的体验或别人的影响所形成的一种看法体系。英语学习者的观念大致分为两类：一类为管理观念，另一类为语言学习观念。管理观念指学生对确定目标、制定计划、选择策略、调控策略等一系列管理活动重要性的认识。语言学习观念是指学生对如何才能掌握好语言知识、语言技能和交际能力的主张。语言学习

---

* 由于版面的限制，这里只能登载该项目的部分研究结果，有兴趣者可阅读该项目的结项报告（见文秋芳等，2000）。

观念* 包括：形式操练观念（即认为掌握语言知识对学好外语非常重要）；功能操练观念（即认为在交际中运用语言知识对学好外语非常重要）；依赖母语观念（即认为通过翻译来学习第二语言是捷径）。

策略指学习者为有效学习所采取的措施，这种行为可以是外部活动，也可能是内部活动。策略分为管理策略与语言学习策略。前者与语言学习过程的管理有关，后者与语言学习材料发生直接的关系。管理策略涉及目标的制定、策略的选择、时间的安排、策略有效性的评估和调整、以及情感的控制与调整。与语言学习观念相对应，语言学习策略包括：形式操练策略、功能操练策略和依赖母语策略。

2. 动机、观念和策略之间的关系

在学习者调控系统内部，动机影响观念，观念影响策略的选择。在观念与策略之间，管理观念影响管理策略的使用和选择；语言学习观念影响语言策略的使用和选择。管理观念又影响语言学习观念；管理策略又影响语言学习策略的成效。从这个意义上说，与管理学习过程有关的观念和策略在高一个层面上，与学习语言知识和技能本身直接有关的观念和策略在低一个层面上。

二、研究设计

1. 研究对象

参加该项研究的学生共有 72 名。他们于 1996 年 9 月考进南京大学外国语学院的英语专业，其中女生 57 名，男生 15 名，平均年龄为 18.7 岁。

2. 研究工具

"学习情况调查问卷"是笔者根据撰写博士论文时所设计的学习者因素问卷（Wen，1993）修改而成，并将其翻译成中文。问卷由四部分组成：1）个人简况；2）英语学习动机；3）管理/语言学习观念；4）管理/语言学习策略。除第一部分外，其余都是选择题。与学习动机相关

* 语言学习观念与语言学习策略分类的理论依据是 Stern（1983）提出的 3 大理论争端，详细内容请阅读 "Fundamental concepts of language teaching"。

的题目有6个，其中3个属于深层动机，3个属于表层动机，学生需要从5个选项中挑选1个答案（即从“这个原因对我不重要”到“这个原因对我极其重要”）。有关观念的问卷题共有22个，涉及管理观念、形式操练观念、功能操练观念和依赖母语观念。学生可挑选的答案也有5个，即从“我坚决不同意这个看法”到“我非常同意这个看法”。有关策略的问卷题共有35道题，涉及管理策略、形式操练策略、功能管理策略和依赖母语策略。每个问卷题后仍旧提供五个选项，即从“这个句子完全或几乎完全不适合我的情况”到“这个句子完全或几乎完全适合我的情况”。

3. 数据收集和分析

笔者请一、二年级教授综合英语课的老师和三年级教授阅读课的老师分别在1996年、1997年和1998年的开学初，让学生在课堂上填写了问卷。问卷的数据输入电脑后，运用社会科学的统计软件（SPSS）对数据进行了分析。分析分4步进行：1）用描述统计列出各项变量的频率，对学生未作回答的个别题目用该题目的平均数代替；2）用内部一致性的方法检验各可控因素归类是否能达到统计上的要求（见附录一），其中有12道题因未达到统计要求而被删除；3）用相关分析的方法，检查变量之间的关系；4）用 t – ests 检验学生在大学两年中英语学习动机、观念和策略变化的情况。

三、研究结果与讨论

1. 动机的变化

这一部分将报告学生的英语学习动机（深层动机和表层动机）在三次（刚进大学时，进大学一年后，进大学两年后）调查中，每次的偏爱倾向是什么，这种偏向是否达到统计意义，然后报告纵向比较的结果，即对动机的三次调查结果是否有变化，如有变化，变化的特点和趋势是什么。

1）偏爱倾向

根据表1和表2，1996年学生对两种学习原因的偏爱程度完全一样，而1997年和1998年，学生的深层动机明显强于表层动机，并且差异具有统计意义；1997年差异程度达到了 .02 的显著水平，1998年达到了

.000的水平。

变量名称	深层动机	表层动机
1996年（平均数/方差）	3.34/.68	3.34/.70
1997年（平均数/方差）	3.63/.63	3.41/.62
1998年（平均数/方差）	3.83/.68	3.44/.73

表1：两个动机变量的描述统计结果

	Mean/SD		T-value	P
1996年	深层动机	3.34/.68	-.04	.972
	表层动机	3.34/.70		
1997年	深层动机	3.63/.63	2.37	0.20
	表层动机	3.41/.62		
1998年	深层动机	3.83/.68	3.80	.000
	表层动机	3.44/.73		

表2：*t*-tests：学生的两种学习动机的变化情况

2）变化的趋势和特点

从纵向比较来看，学生的深层动机在逐年加强，且每两次之间的变化都具有统计意义（见表3）。深层动机问卷题的平均数1996年是3.34，1997年增加到3.63、1998年增加到3.84。根据表4，表层动机也在逐年加强，1996年表层动机问卷题的平均数为3.34，1997年增加到3.41，1998年增加到3.44。与深层动机不同的是，表层动机每两次之间的变化都不具有统计意义。

深层动机	Mean/SD	相关系数	P	T-value	P
第一次与第二次比较（1996年/1997年）	3.34/.68 3.63/.63	.65	.000	-4.57	.000
第二次与第三次比较（1997年/1998年）	3.63/.63 3.83/.68	.56	.000	-2.66	.010

表3：*t*-tests：1996-1998年深层动机变化的情况

表层动机	Mean/SD	相关系数	P	T - value	P
第一次与第二次比较（1996年/1997年）	3.34/.70	.50	.000	-.89	.376
	3.41/.62				
第二次与第三次比较（1997年/1998年）	3.41/.62	.41	.000	-.27	.791
	3.44/.73				

表 4：*t* - tests：1996 - 1998 年表层动机变化的情况

2. 观念的变化

1）偏爱倾向

从 1996 年到 1998 年，在 4 种不同的观念中，管理观念的平均数最高，依赖母语观念的平均数最低、功能操练观念的平均数位于第二，形式操练观念的平均数位于第三（见表 5）。虽然功能观念的平均数高于形式操练的平均数，但只有 1998 年的差异具有统计意义。

变量名称	管理观念	形式操练观念	功能操练观念	依赖母语观念
1996 年(平均数/方差)	4.25/.39	3.89/.40	3.91/.36	2.66/.55
1997 年(平均数/方差)	4.24/.44	3.88/.40	3.94/.40	2.70/.43
1998 年(平均数/方差)	4.30/.45	3.81/.46	3.91/.41	2.73/.49

表 5：不同观念的平均数和标准差

（注：这里每种观念的平均数 = 每个问卷题的平均数相加后再除以题目数。）

2）变化的趋势和特点

从纵向的角度来比较，4 种观念的平均数虽有或多或少的变化，但它们都不具有统计意义。这进一步证明了 Wenden（1991）提出的观念稳定性，即观念一旦形成，它和其他知识一样储存在长期记忆中，成为知识体系的一部分，不容易改变。或者说，学生在进入大学时，调查中所涉及到的 4 类观念已经趋向于稳定。

3. 策略的变化

1) 偏爱倾向

横向比较1996年到1998年的各类策略平均数（见表6）。除1998年形式操练策略和功能操练策略之间没有显著性差异以外，其他各类策略之间的差异都具有统计意义。在4项策略中，管理策略的平均数最高，母语策略最低，形式操练策略和功能操练策略的平均数处于中间，这一总体趋势与观念相同。所不同的是，功能操练观念的平均数每次都高于形式操练观念的平均数，（1996年，.002；1997年，.006；1998年，.10）；而功能操练策略的平均数在1996年和1997年显著低于形式操练策略（1996年，－.42；1997年，－.24），只有到了1998年情况才发生了变化，功能操练策略的平均数稍许高出了形式操练策略（1998年，.02），但差异不具有统计意义。

由此，我们可以得出结论：学生最偏爱管理策略，最不喜欢母语策略；在刚进大学和经过一年学习之后，喜欢形式操练策略的程度明显强于喜欢功能操练策略的程度；但经过第二年的学习以后，学生对形式操练和功能操练两类策略的使用不再有明显的偏爱。许多西方的研究者们把形式操练和功能操练两类策略看成是互为对立的一对矛盾，认为只有使用后一种策略，学习者的外语交际能力才能得到发展。而我校学生在两年内使用这两类策略总趋势的变化似乎表明它们在实际学习的过程中并不互相对立。

变量名称	管理策略	形式操练策略	功能操练策略	母语策略
1996年（平均数/方差）	3.55/.48	3.39/.45	2.97/.52	2.66/.77
1997年（平均数/方差）	3.52/.47	3.40/.43	3.16/.49	2.57/.60
1998年（平均数/方差）	3.61/.52	3.29/.48	3.31/.51	2.43/.62

表6：1996－1998年4种策略的平均数及方差

（注：这里每种策略的平均数＝每个问卷题的平均数相加后再除以题目数）

2) 变化的趋势与特点

如表7所示，1996年到1998年，管理策略的平均数虽有或多或少的变化，但都不具有统计意义。换句话说，学生的管理策略没有受到大学

两年生活的影响。或者说，学生的管理策略在高中毕业时已趋向成熟。

管理策略	Mean/SD	相关系数	P	T－value	P
第一次与第二次比较（1996年/1997年）	3.55/.48 3.52/.47	.57	.000	－.70	.486
第二次与第三次比较（1997年/1998年）	3.52/.47 3.61/.52	.56	.000	－1.54	.128

表7：*t*-tests：1996-1998管理策略的变化情况

根据表8，形式操练策略的变化总趋势是先升后降，第一次变化，即上升，没有统计意义，但第二次变化，即平均数降低了.11，达到了.028显著性水平。这似乎表明形式操练策略变化的关键期在二年级，而不在一年级。这一总体趋势与笔者1996年跟踪调查的结果相同，所不同的是，在1996年的跟踪调查中，两次变化都没有统计意义。

形式操练策略	Mean/SD	相关系数	P	T－value	P
第一次与第二次比较（1996年/1997年）	3.39/.45 3.40/.43	.35	.003	－.14	.889
第二次与第三次比较（1997年/1998年）	3.40/.43 3.29/.48	.63	.000	2.25	.028

表8：*t*-tests：1996-1998形式操练策略的变化情况

与形式操练策略不同，功能操练策略的平均数呈直线上升趋势，每次变化都具有统计意义（见表9）。这表明了学生通过大学两年的英语学习，使用功能操练策略有了急剧的增加。这一结果并不出乎意料。英语专业的教学方法和要求都与中学有明显的不同，学生必须要具备听说读写4方面的交际能力。为达到这一要求，学生一定要增加运用语言进行真实交际的活动。这一结果与笔者1996年的跟踪调查结果基本相同，所不同的是，本项目两次变化都具有统计意义，而1996年的结果只有第二次变化才具有统计意义。

功能操练策略	Mean/SD	相关系数	P	T－value	P
第一次与第二次比较（1996年/1997年）	2.97/.52 3.16/.49	.40	.001	－2.99	.004
第二次与第三次比较（1997年/1998年）	3.16/.49 3.31/.51	.65	.000	－2.89	.005

表9：*t*-tests：1996-1998年功能操练策略的变化情况

与功能操练策略相反，依赖母语策略的平均数呈直线下降的趋势，且每两次的变化都具有统计意义（见表10）。这一结果比较符合逻辑。对母语的依赖程度往往是检验外语水平的标尺。随着外语水平的提高，学生对母语的依赖程度应该逐步降低。以上结果与1996年的跟踪调查结果有相似之处。1996年的结果是：先降后微升，只有第一次变化具有统计意义。

依赖母语策略	Mean/SD	相关系数	P	T－value	P
第一次与第二次比较（1996年/1997年）	2.66/.77 2.57/.60	.48	.000	－2.99	.004
第二次与第三次比较（1997年/1998年）	2.57/.60 2.43/.62	.51	.000	－2.89	.005

表10：*t*-tests：1996-1998年母语策略的变化情况

4. 三类可控因素之间的关系及其变化

1）动机与观念和策略

表11显示表层动机与4种观念之间的相关系数均没有统计意义，但与管理观念的正相关有明显增强的趋势。深层动机与管理观念和形式操练观念分别呈正相关，其中97年深层动机与管理观念的相关系数.30达到.009的显著水平，96年深层动机与形式操练观念的相关系数.25达到.032的显著水平；与依赖母语观念呈负相关，且有增强的趋势，其中1997年和1998年的相关系数（－.24和－.27）均有统计意义，显著性水平分别为.05和.02；与功能操练观念没有联系。

表层动机与管理策略和功能操练策略呈正相关，其中1997年，1998年表层动机与管理策略的相关系数为.29和.47，分别达到.015和.000的显著性水平，1997年表层动机与功能操练策略的相关系数为.26，达

到 .029 的显著性水平；与其他策略没有统计意义上的联系。深层动机与依赖母语策略的关系最为紧密，1996 年，1997 年和 1998 年 3 次数据均显示它们之间的关系是负相关（ - .41， - .31， - .38），且每次相关的程度都达到 .001 以上的显著性水平。深层动机与管理策略的关系密切程度仅次于依赖母语策略，1997 年与 1998 年它们之间的相关系数为 .46 和 .34，显著性水平分别为 .000 和 .004。深层动机与功能操练策略的关系也呈正相关，1997 年和 1998 年两次的相关系数（.26 和 .23）分别达到 .026 和 .049 的显著性水平。在四类策略中，惟一与深层动机没有关系的策略就是形式操练策略。

表层动机	管理观念/策略	形式操练观念/策略	功能操练观念/策略	依赖母语观念/策略
1996 年相关系数	.01/.16	.10/.15	.17/.09	.08/.09
1997 年相关系数	.11/.29 *	.19/.17	.05/ - .12	.01/.08
1998 年相关系数	.21/.47 * * *	.18/.08	.11/.26 *	.10/ - .03
深层动机	管理观念/策略	形式操练观念/策略	功能操练观念/策略	依赖母语观念/策略
1996 年相关系数	.22/.13	* .25/ - .06	.04/.15	- .13/ - .41 * * *
1997 年相关系数	* * .30/.46 * * *	.07/.21	.01/.26 *	* - .24/ - .31 * * *
1998 年相关系数	.21/.34 * *	.14/.08	.10/.23 *	* - .27/ - .38 * * *

表 11：动机与观念和策略的关系

（注： * < .05， * * < .01， * * * < .001）

与动机和观念的关系相比，动机和策略的关系更为密切。前两者之间有统计意义的相关系数只有 4 个，后两者有统计意义的相关系数有 10 个。而在这 14 个有统计意义的相关系数中，深层动机占有 11 个，而表层动机只占 3 个。这意味着深层动机对学习者的观念与策略影响大于表层动机。

在各类观念和策略中，管理观念/策略与动机有统计意义的相关系数占有 5 个，依赖母语观念/策略与动机的也占有 5 个，其余 3 个在功能策略中，1 个在形式操练观念中。这意味着动机对管理观念/策略，对依赖母语观念/策略的影响比较大，其次是功能策略，最弱的是形式操练观念；对功能操练观念与形式操练策略几乎没有影响。

在这些相关系数中，特别值得讨论的有 3 点。第一点是深层动机与

依赖母语观念/策略的关系。在它们所有的6对关系中，都呈负相关，其中5对具有统计意义。这说明它们之间的关系方向一致，并且比较稳定。这一结果在以前类似研究中从未发现过，但比较符合逻辑。具有深层动机的学生对学习外语感兴趣，对外国文化感兴趣，对外国的科学技术感兴趣。深层动机愈强，使用目标语言的欲望愈强，因此依赖母语的可能性就越小。

第二个值得讨论的是两类动机与管理策略的关系。在6对关系中（见表11），所有都呈正相关，其中后两年（即1997年和1998年）4对关系具有统计意义。这一结果意味着学生在刚进大学时，动机对管理策略的影响还不明显，但通过大学第一年的学习以后，动机与管理策略之间的关系有了显著的加强。无论是表层动机，还是深层动机，只要想学英语，就都会对管理策略产生积极的影响。

第三个值得讨论的是动机与功能操练策略的关系。6对关系中，5对正相关，其中三对具有统计意义。1996年，当学生刚进入大学时，两种动机与功能操练策略的关系均不具有统计意义。1997年，当学生进入大学一年后，深层动机与功能操练策略的相关系数为.26，达到了.026的显著性水平。到了1998年，即第三次调查时，表层动机和深层动机与功能操练策略的相关系数分别为.26和.23，均达到.05的显著性水平。这似乎意味着通过大学两年的学习，动机与功能操练策略的关系在逐步建立。

从总体上看，相对于其他策略而言，动机与形式操练策略关系最不密切。出现这一结果的原因可能是，与其他几类策略相比，学生对形式操练策略最为熟悉，实践最多，同学之间的差异最不显著。在各类观念中，两类动机与功能操练观念关系最不密切，其次是形式操练观念。出现这一结果并不令人奇怪。观念只是学习者头脑中的看法，而不是学习者自身的行为。看法对行为虽有一定的影响，但两者之间一定会有差距，特别是当观念的实现需要学习者克服困难时，这种差距更为明显。在这种情况下，比较普遍的现象是：想得好，做得差。

2）观念与策略

从1996年到1998年，管理观念/策略、形式操练观念/策略、功能操练观念/策略与依赖母语观念/策略都呈现正相关（见表12）。在12对相关系数中，除1996年形式操练观念/策略，1996年与1997年功能操练观念/策略三对以外，其他各对都达到了显著性水平。从总体上看，四

对观念/策略的关系都有逐年增强的趋势。这意味着学生的某种观念越强，使用相对应的策略可能性越大。出现这一结果的主要原因有两个。第一个原因是心理因素。随着学生的年龄增加，他们逐年成熟起来，观念与策略之间的差距也应该随之缩小。第二个原因与英语水平有关系，即随着学生英语水平的提高，原来比较难以实现的观念现在逐渐变得容易付诸实践。

变量名称	管理	P	形式操练	P	功能操练	P	依赖母语	P
1996 年相关系数	.27	.02	.12	.28	.20	.09	.38	.00
1997 年相关系数	.36	.00	.29	.01	.17	.16	.40	.00
1998 年相关系数	.56	.00	.39	.00	.43	.00	.48	.00

表 12：观念与策略之间的关系

在这 4 对观念与策略的关系中，其中依赖母语观念与依赖母语策略的关系最为稳定，1996 年，1997 年和 1998 年 3 次的相关系数分别为 .38，.40 和 .48，差异最小。经过大学第一年的学习后，形式操练观念与形式操练策略之间的关系发生了最为剧烈的变化，1996 年与 1997 年的相关系数分别为 .12 和 .29，相差 .17；其次是管理策略与管理观念之间的关系，1996 年与 1997 年的相关系数分别为 .27 和 .36，相差 .09；其余两对观念/策略相关系数的差别只有 .03 和 .02。经过大学两年学习后，除形式操练观念/策略以外，3 对观念/策略之间的关系发生的变化都远远大于第一年，1996 年与 1997 年的差异分别是：管理观念/策略 .09、功能操练观念/策略 .03，依赖母语观念/策略 .02；而 1997 年与 1998 年的差异分别是：.20，.26，.08 。这似乎说明观念/策略之间关系密切程度的变化，关键期在大学学习的第二年。

四、结论

1. 主要结果

现将该项目的主要研究结果总结如下：

1）学生在刚进大学时，对表层和深层动机没有明显的偏爱倾向，经过大学第一年学习后，深层动机的强度超过了表层动机；经过大学第二年学习后，深层动机成为明显强势动力源。尽管两种动机的强度都是

逐年增强，但深层动机强度的增加幅度远远超过表层动机。两种动机之间的关系由负相关转为了正相关。在4种观念中，学生最赞成管理观念，最不赞成依赖母语观念，对形式操练观念和功能操练观念的赞成程度位于两者之间。这一倾向经过两年的学习没有变化。在四种策略中，学生使用管理策略最多，使用依赖母语策略最少，对形式操练策略和功能操练策略的使用程度位于两者之间。这一总体倾向经过两年的学习没有变化，但就形式操练策略和功能操练策略而言，学生的偏爱倾向有明显变化。1996年刚进大学时，学生使用形式操练策略的程度显著高于使用功能操练策略；1997年时，虽然学生使用功能操练策略的程度明显增加，但仍旧低于形式操练策略的使用程度；1998年时，使用功能操练策略的程度继续增加，而使用形式操练策略的程度反而有所下降，因此使用功能操练策略的程度高于使用形式操练策略的程度。

2）经过大学两年学习，深层动机和功能操练策略的使用程度呈逐年上升趋势，且具有统计意义；依赖母语策略的使用程度呈逐年下降趋势，且具有统计意义；形式操练策略的使用程度先微升后下降，微升不具有统计意义，但下降具有统计意义；所有四种观念以及管理策略均无显著变化。深层动机、功能操练策略和依赖母语策略的变化属于渐进式，即经过大学第一年和大学第二年的学习，它们都发生了具有统计意义的变化；形式操练策略的变化属于突变式，即第一年没有发生具有统计意义的变化，到了第二年才发生这样的变化。

3）根据3次跟踪调查的数据，尽管代表各类可控因素的变量可呈上升或下降趋势，各类因素之间关系的密切程度可增强或减弱，但各类因素之间关系的趋向比较稳定（见图1）。深层动机与管理观念/策略、功能操练策略以及依赖母语观念/策略有联系；表层动机与管理策略有联系；4个观念与相应的4个策略有联系；管理观念与形式操练观念以及功能操练观念有联系；管理策略与形式操练策略以及功能操练策略有联系。值得指出的是：管理观念/策略、形式操练观念/策略和功能操练观念/策略与依赖母语观念/策略没有联系。惟一与母语观念/策略有联系的是深层动机。

4）动机与观念以及策略的关系密切程度的变化没有明显的规律可循，但策略与观念的关系有逐年加强的趋势，其关系密切程度的变化关键期主要在大学学习的第二年。

以上各项结果需要在今后的研究中进一步验证。

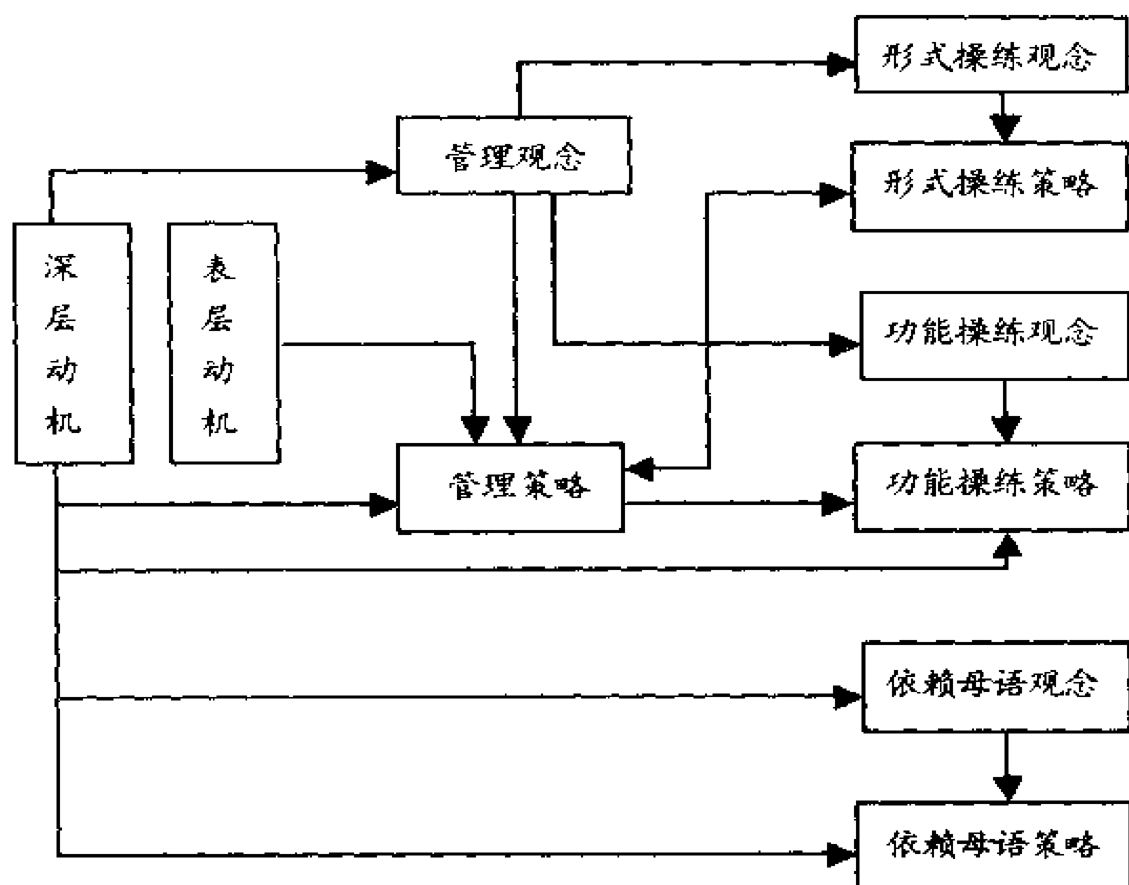


图 1：本项目中各变量之间的关系

2. 该项研究结果的意义

本项目的研究结果表明动机、观念、策略之间的关系具有较高的稳定性。动机影响观念和策略；观念影响策略；管理观念/策略影响形式操练策略/功能操练策略，但不影响依赖母语观念/策略。因此在教学中，教师如果要改变学生的语言学习行为，必须依据这些因素之间的关系，从总体上考虑，研究调控措施，避免简单从事。

深层动机、功能操练策略以及依赖母语策略在大学第一年和第二年以后都发生了显著性的变化；形式操练策略的变化主要发生在第二年以后。这一结果说明英语专业的第一年和第二年是学生可控因素发生变化的关键期。如果在这个阶段能够配备得力的教师，对学生因势利导，一定能收到事半功倍的效果。

3. 对未来研究的建议

1）该项目的不足之处

该项目的不足之处包括以下几个方面：

（1）被调查的学生人数少，且限于某一个学校，因此代表性不够广泛。

（2）定性调查由于设计不够周密，因此数据不能用于进一步分析。

（3）同一问卷调查了3次，学生可能产生厌烦情绪，影响数据的可靠性。

鉴于以上不足，在应用本项目的研究结果时需要格外谨慎。

2）未来研究可控因素变化的几种可能性

像这样动态跟踪研究可控因素变化的项目，在该领域内尚属首次。我们认为可以选用相关的样本来重复该项研究，检验本次研究的结果。此外，该项目还需要从以下几个方面继续深入研究：（1）运用定性的方法跟踪研究少数学生的可控因素变化情况，当然由于定性研究费时费力，事先一定要精心策划，收集数据时一定要仔细认真；（2）在本次研究的基础上着重深入研究依赖母语观念/策略的变化与深层动机的关系；（3）深入研究某些可控因素的变化与教学过程之间的关系。

参考文献

Biggs, J.B.(1979). Individual differences in study processes and the quality of learning outcomes. Higher Education, 8, 381-394.

Cohen, A.(1998). Strategies in learning and using a second language. London: Longman.

Ellis, R.(1994). The study of second language acquisition. Oxford: Oxford University Press.

Gardner, R. & Lambert, W.(1972). Attitudes and motivation in second-language learning. Rowley, Mass.: Newbury House.

Huang, X. & M. Van Naerssen.(1985). Learning strategies for oral communication. Applied Linguistics, 8(3), 287-307.

Oxford. R.(1990). Language learning strategies: What every teacher should know. Rowley, Mass.: Newbury House.

Stern, H. (1983). Fundamental concepts of language teaching. Oxford: Oxford

University Press.

Wen, Qiufang. (1993). Advanced level English learning in China: The relationship of modifiable learner variables to learning outcomes. Hong Kong University: Unpublished doctoral thesis.

Wen, Qiufang & Johnson, R.K. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. Applied Linguistics, 18(1), 27-48.

Wenden, A. (1991). Learner strategies for learner autonomy. New York: Prentice Hall.

王文宇，1998，观念、策略与词汇记忆，《外语教学与研究》第 1 期。

文秋芳，1995，英语成功者与不成功者在学习方法上的差异，《外语教学与研究》第 3 期。

文秋芳，1996，大学生英语学习策略变化的趋势及其特点，《外语与外语教学》第 4 期。

文秋芳、王海啸，1996a，大学生英语学习观念与策略的分析，《解放军外语学院学报》第 4 期。

文秋芳、王海啸，1996b，学习者因素与大学英语四级考试成绩的关系，《外语教学与研究》第 4 期。

文秋芳、乐眉云、丁言仁、王文宇，2000，九五社科项目结项报告：英语学习者动机、观念、策略的变化规律与特点。

附录

动机名称	问卷题内容	题目数	Alpha (1996 年)	Alpha (1997 年)	Alpha (1998 年)
深层动机	1. 我喜欢学习外国语 2. 我对说英语国家的文化感兴趣 3. 我要了解外国的科学技术、政治和经济情况	3	.33	.42	.62
表层动机	1. 我要找一份好工作 2. 我要报考研究生 3. 我要出国深造	3	.41	.47	.54

表 1：两类动机的问卷题内容、题目数量和 Alpha 值

观念名称	问卷题内容	题目数	Alpha 1996年	Alpha 1997年	Alpha 1998年
管理观念	1. 很好地计划自己的学习时间是学好英语的重要保证 2. 选择有效的学习方法对学好英语很重要 3. 有明确的长期和短期学习目标对学好英语很重要 4. 不断总结自己的进步并找出存在的问题对学好英语很重要 5. 学好英语的一个重要方面是了解自己的个性特点，发挥长处，克服短处 6. 经常反思自己的学习方法是否有效，对学习外语非常重要	6	.52	.74	.76
形式操练观念	1. 记外语句型，对学习外语非常重要 2. 记单词对学好外语很重要 3. 精读课文对学习外语很重要 4. 背诵好的课文对学习外语很重要 5. 反复朗读课文对学好外语很重要 6. 反复模仿好的录音带对练好语音语调很重要	6	.46	.62	.70
功能操练观念	1. 进行大量的听力练习对学好外语很重要 2. 要学好外语，大量阅读外文报纸、杂志、小说等很重要 3. 猜单词和句子的意思是解决阅读困难的最好方法 4. 进行大量的口语训练，对学好外语很重要 5. 进行大量的写作练习，对学习外语很重要 6. 听英语时，碰到没有学过的单词，最好的方法是猜测	6	.42	.62	.60

（见下页）

（接上页）

观念名称	问卷题内容	题目数	Alpha 1996年	Alpha 1997年	Alpha 1998年
母语观念	1. 理解课文最好的方法是翻译 2. 说英语时，最好用中文想好要说的内容 3. 要想将所听的英文内容记下来，最好的方法是记中文意思 4. 要想写出好的英语作文，最好的方法是先用中文组织好想写的内容	4	.59	.69	.79

表2：4种观念的问卷题内容、题目数量和Alpha值

策略名称	问卷题内容	题目数	Alpha 1996年	Alpha 1997年	Alpha 1998年
管理策略	1. 除了老师布置的作业外，我有自己的学习计划。 2. 为了使自己有足够的时间学习英语，我很好地安排自己的学习日程。 3. 我对改进自己的英语学习有明确的要求 4. 我评价自己学习英语进步的情况，从而找出薄弱环节和改进措施 5. 我评价自己的学习策略，从而找出存在的问题和解决的方法。 6. 我根据学习任务的特点，选择不同的学习策略。 7. 我选择适合自己英语水平的材料来学习。 8. 我研究自己的个性特点，找出哪些特点有利于自己的英语学习，哪些特点阻碍自己的进步，从而能发挥自己的优势，采取相应的措施，克服弱点。	8	.74	.78	.84

（见下页）

（接上页）

策略名称	问卷题内容	题目数	Alpha 1996年	Alpha 1997年	Alpha 1998年
形式操练策略	1. 当阅读课文时，我争取弄懂课文里的每一处 2. 课外，我反复朗读课文 3. 当我不懂句子的意思时，我分析句子的语法结构 4. 我背诵好的课文 5. 当听英语材料时，我争取听懂每一句话的意思 6. 假如在听英语时碰到生字，我会尽量记住生字的发音，然后根据发音在字典上查找它的拼写和意思	6	.49	.61	.62
功能操练策略	1. 我在课外主动阅读英文报纸、杂志和小说 2. 我课外主动听英语广播。 3. 在英语课上，我尽量主动地争取回答问题 4. 我课外主动看英语电视和电影。 5. 我课外尽量用英语与同学和老师会话 6. 我课外自己对自己说英语 7. 我主动地用英语记笔记、留言、写信或日记。	7	.60	.71	.73
母语策略	1. 为了帮助对课文的理解，我把课文译成中文 2. 我说英语时，首先用中文组织意思，再翻译成英语。 3. 我用英文写作文时，先用中文组织意思，再译成英语。 4. 听英语时，我用中文记住所听的意思。	4	.82	.72	.79

表 3：4 种策略的问卷题内容、题目数量和 Alpha 值

附录二：实验研究

输入方式与听力词汇习得——一项听力词汇习得实验的报告

南京农业大学 王艳

提要：听力词汇少是造成学生听力理解困难的重要原因。本论文报告了不同输入方式对听力词汇习得影响的实验。结果发现，声音输入组在生词音的辨识和意义的理解上都强于字形输入组，即听入比读入更有利于听力词汇的习得。结果还发现，阅读词汇通过听力练习可以转化为听力词汇。

关键词：听力词汇、阅读词汇、词汇输入方式

一、引言

笔者在多年听力教学中，通常听到学生反映生词是他们听力理解的主要困难，尤其是大学低年级学生。1999 年笔者曾以“英语听力理解困难”为题对大学一年级学生进行了调查。在学生自我报告的困难中，词汇困难占第一位。刘思（1995）在研究中也发现学生的听力词汇量明显低于阅读词汇量，由此可以推断提高听力词汇量是提高听力理解能力的关键之一。

Nation（1990）认为词汇习得的方式是影响词汇学习的重要因素之一，即单词的初始输入是通过听觉渠道还是通过视觉渠道。国外在这方面的实证研究不多。Channell（1988）认为视觉渠道不是学习生词的最好输入方式。Feitelson, Goldstein, Iraqi & Share（1993）发现，通过听觉渠道来输入生词不仅易于学会单词，而且能帮助学习语法。国内对词汇输入方式与单词习得关系的研究几乎是空白。本文报告的实验只是对这一问题研究的初步尝试。笔者希望通过实验回答两个问题：（1）听入是否比读入更有利于听力词汇的习得？和（2）阅读词汇可否转化为听力词汇？

二、研究设计

1. 受试对象

该研究的受试对象为南京农业大学人文学院英语专业 2000 年进校的

一年级全体学生。他们来自全国15个省份中的40多个不同城市和3个直辖市，共56人，其中11名男生，45名女生，平均年龄为18.9岁，高考英语平均成绩为112分。

这些学生在进校时随机地分成两个班，每班均为28人。为确认两班学生的英语听力水平相当，实验前进行了一次听力测试，内容为3篇短文，10道选择题，1题1分。测试内容选自外语教学与研究出版社1996年出版的《初级英语听力》。表1为测试结果。

	平均数	标准差
1班（N=28）=实验组	7.4286	1.451
2班（N=28）=对比组	7.5357	1.503

表1：实验前测试结果

经 t 检验，两班成绩没有显著性差异。1班作为实验组，采用听人的方式学习生词（以下称为听人组）；2班作为对比组，采用读人的方式学习生词（以下称为读人组）。

2. 实验设计与步骤

实验组与对比组，按不同的输人方式学习10个生词。其中，5个为目标生词，即这些词将会在测试的短文中出现；5个为干扰词，从字典上随机选出（见附录1）。10个词混放在一起，受试者需要判断这些词是否在短文中出现。实验分为两个阶段进行：1）词汇学习；2）测试。

两组受试者在第一阶段采用不同的词汇学习方式：

1）听人组：通过听的方式学习给定的10个单词，即教师将每个单词念3遍，每次的朗读速度和中间的停顿都保持一致，并用汉语给出中文意思一遍。平均每个词需要9秒钟时间，总时间约为1.5分钟。10个单词给出的次序事先随机决定。

2）读人组：采用阅读的方式学习10个生词。教师将单词及其中文意思写在胶片上，用投影仪投映在银幕上。10个单词给出的次序与听人组相同，总时间约为1.5分钟。

第二阶段为测试，共进行两次。两组采用的测试方式完全相同。第一次测试时，先播放一篇短文录音（见附录2），然后教师将10个单词依次念出，让受试者判断哪些词在这篇录音材料出现过。第一次测试结

果交上来之后，进行第二次测试。第二次测试内容分为两个部分。第一部分和第一次测试内容相同，第二部分为新增加的听力理解题，即受试者根据对短文的理解从三个选项中选出一个正确答案，问题和选项均由教师念出。第一部分测试内容的目的是比较两组受试者在掌握目标词的音上是否有差异，第二部分内容旨在通过学生对短文理解的程度来推测两组学生对生词意义的掌握是否有差异。考虑到受试者是一年级学生，其整体听力水平比较低，因此听力理解题安排在第二次测试中。这样，两组受试者在听两遍短文的基础上再做选择题，可降低因随意猜测答案而带来误差的可能性。短文共 157 个词，掌握 5 个目标词的意义对短文理解都很关键。表 2 列出了实验步骤和所需时间。

次序	步骤		目的	时间(分钟)
	实验组(听入)	控制组(读入)		
1	实验前测试	实验前测试		20
2	学习生词(听入)	学习生词(读入)	显示不同的输入形式	1.5
3	实验后第一次测试(听短文，做词汇测试)	实验后第一次测试(听短文，做词汇测试)	测试学生通过听短文来辨认已学过单词的能力	5.0
4	实验后第二次测试(听相同短文，做词汇与听力理解测试)	实验后第二次测试(听相同短文，做词汇与听力理解测试)	1. 测试增加听力训练的次数是否提高辨认单词的准确性 2. 用听力理解题推测学生对生词词义的理解情况	5.0

表 2：实验步骤、目的和所需时间

3. 数据收集

两组受试者在听完一遍短文之后，在纸上写下 1 到 10，以代表单词的顺序。教师将 10 个单词读出来，让受试者作出判断。如果他们认为某个词刚才在短文中出现过，就在相应的题号后打√，如果不是，就打×。第二次测试中，受试者以同样的方式，做一遍词汇测试，然后做短文的听力理解题。受试者根据自己的理解，选出认为正确的选项。

4. 数据分析

在分析词汇测试的结果时，受试者每正确判断一个词得1分，即对确实从短文中挑选出来的目标词打√的给一分，和对干扰词打×的也给1分。凡是错的均不减分。满分为10分。听力理解选择题选对的得1分，选错的不减分，全对得3分。所有数据输入计算机，用SPSS进行处理。笔者计算了每个目标单词的平均分和10个单词的总平均分，以及理解题的平均分（见表3），然后将两组的词汇成绩和听力理解成绩进行了独立样本T检验，最后将听入组和读入组的第一次词汇测试成绩与第二次的分别进行配对样本T检验。

	第一次测试		第二次测试	
项目	平均数	标准差	平均数	标准差
听入组（N=28）				
目标词汇				
Rhythm	0.54	0.51	0.54	0.51
Crawl	0.96	0.19	0.96	0.19
Spur	0.71	0.46	0.61	0.50
Dismay	0.68	0.48	0.68	0.48
Stroke	0.57	0.50	0.61	0.50
10个目标词总分	6.64	1.39	6.82	1.90
段落理解题得分			2.60	0.57
读入组（N=28）				
目标词汇				
Rhythm	0.50	0.51	0.39	0.50
Crawl	0.68	0.48	0.82	0.39
Spur	0.54	0.51	0.68	0.48
Dismay	0.64	0.49	0.75	0.44
Stroke	0.61	0.50	0.82	0.39
10个目标词总分	5.64	1.70	6.50	1.64
段落理解题得分			2.25	0.65

表3：测试结果

三、结果与讨论

1. 听入是否比读入更有利于听力词汇的习得?

词汇的音和义是掌握听力词汇最基本的两个方面。根据表4，在第一次测试中，听入组的词汇平均分（6.64）明显高于读入组（5.64）。经独立样本T检验，两组的得分具有显著性差异（$t = 2.40$，$p < .05$）。

组别	平均数	标准差	p
听入组（N = 28）	6.64	1.39	.02
读入组（N = 28）	5.64	1.70	

表4：两组在第一次测试中的总分比较

这说明，第一次测试中听入组在单词音的记忆和辨别上强于读入组。可以推断，采用听入的方法学习词汇，能够较好地识别单词的发音。这一结果符合心理学的理论。两种学习方法接受的外界刺激不同，因此在大脑中形成的感觉记忆表象也不同。在外界刺激上，一个是声音，一个是字形。根据记忆的一般原理，人脑对信息的存储是以感觉痕迹形式被记录下来。通过听觉接受的输入，记录的就是声音表象；通过视觉接受的输入，记忆的就是图像。心理学家认为声像比图像的延续时间长（朱纯，1994：188）。当原始刺激再度呈现时，受试者需要再认知，即将当前呈现的刺激与头脑中记忆表象相匹配。很显然，当受试者听含有目标词汇的短文录音材料时，听入组的记忆痕迹与听到的目标词更加接近，因此识别任务的完成比读入组要容易得多。由此可以推断，仅仅通过读入方式学习，不能有效地记忆单词的发音，在以后再听到该词时，也不能很好识别。

从表3中可以发现，5个目标单词中，“crawl”这个词的两组得分差距特别大，其差异具有统计上的显著意义（$t = 2.95$，$p < .01$）（见表5）。

词汇	听入组（$N = 28$）		读入组（$N = 28$）		P
	平均数	标准差	平均数	标准差	
Crawl	0.96	0.19	0.68	0.48	.006

表5：第一次测试中两个组对目标词汇“crawl”的得分情况和比较

短文中，"crawl"一词出现的频率为3次，而其他词只出现一次。虽然两个组第一次测试中都只听一遍短文，词汇测试的形式也一样，但是这个词听入组得了0.96分，明显高于其他词；而读入组只得了0.68分，与其他单词的得分相当。这说明目标词出现频率对听入组的影响比对读入组的要大。听入组因依赖其听觉去记忆和辨别单词，听力材料中单词的重复频率对其产生了作用，出现次数越多，越容易识别，因此绝大部分人都判断正确了（全班都答对为1分）。然而，读入组依赖视觉来记忆单词，因此听力材料中单词的重复频率并没有对其产生明显作用。这也证明用读入方法学习词汇不如听入更有利于对词音的辨识。

另外，听入组在对干扰词的判断上，也强于读入组。在5个干扰词中，听入组误选的平均为1.8个，而读入组为平均2.3个。

表6列出了两组在听力理解成绩上的差异。听入组的得分明显高于读入组，且达到统计上的显著意义（$t=2.20$，$p<.05$）。

组别	平均数	标准差	P值
听入组（$N=28$）	2.60	.567	.032
读入组（$N=28$）	2.25	.645	

表6：两组听力理解测试的得分情况

由于5个目标词均为理解短文的关键词，因此我们可以推断，听力理解题得分高的听入组就意味着对目标词词义理解也比读入组好。换句话说，听入组不仅仅是在识别发音上占有优势，在习得词义上也好于读入组。这一结果说明掌握听力词汇不仅可以识别含有生词的语言信息（语音），还有助于对词汇、句子，甚至段落的意义理解。因而，综合前一点的结果，相比较而言，采用听入的方式更有利于习得听力词汇。

2. 阅读词汇可否转化为听力词汇？

表7显示两个组第二次测试的结果已趋于接近。这一次，听入组得分的平均值虽然仍高于读入组，但是两组间得分已不再具有显著性差异。

组别	平均数	标准差	P值
听入组（$N=28$）	6.82	1.90	
阅读组（$N=28$）	6.50	1.64	*n.s.*

表7：两组在第二次试后测试的得分比较

为了进一步了解第二次听短文对受试者的影响，笔者进行了听前与听后的配对检测（a paired-samples t-test）。结果显示，第二次听短文对读入组受试者的影响颇为明显，（$t = -2.25$，$p < .05$），而对听入组受试者的影响并不显著（见表 8）。

组别	听前		听后		P值
	平均数	标准差	平均数	标准差	
听入组（$N = 28$）	6.64	1.4	6.82	1.9	*n. s.*
读入组（$N = 28$）	5.64	1.7	6.50	1.6	.033

表 8：听第二次段落对两个组的影响

这说明，听入组再接受一次相同输入，对听力词汇习得成绩的提高没有明显作用；但对读入组有显著作用。从这一结果我们可以推断，只要增加听力训练，阅读词汇可以转化为听力词汇。

四、结论

本次实验得到了以下结果：

1. 听入比读入更有利于听力词汇的习得。

2. 通过训练，阅读词汇可以转化为听力词汇。

事实上，在词汇学习中，言语听觉和视觉并不是孤立存在的，它们和人的其他感觉器官一起发挥作用。问题在于，听觉和视觉分别是听力和阅读的感性基础，在训练不同的英语技能时应有所侧重。长期以来，我们一直重阅读、轻听说，造成听力词汇量远远不足。真正习得一个单词意味着对其书面及口头形式的理解和使用。而观察学生平时的词汇学习不难发现，他们通常的学习方式与读入组非常相似。他们熟谙传统的读背方式，将单词表中的生词朗读出来，记住拼写和中文意思，却常常忽略了发音。下一次看到这个词时，一时难以识别，误以为是生词。如果我们强调用听觉记忆来记住单词，培养学生先记住音和义、再记住拼写的习惯，就会有助于扩大听力词汇量，提高听力理解能力。如果同时配合视觉和动觉等多种输入方式，听力词汇量和阅读词汇量之间的差距就能缩小。

由于该项实验仅是探索性研究，且受试者人数少，测试只涉及 10 个目标词汇和一篇短文，实验时间也比较短，因此以上结果有待进一步验证。

参考文献

Channell, Joanna. (1988). Psycholinguistic considerations in the study of l2 vocabulary acquisition. In R. Carter & M. McCarthy (Eds.). *Vocabulary and language teaching*, 83-96. London: Longman.

Feitelson, D., Z. Goldstein, J. Iraqi, and D. Share. (1993). Effects of listening to story reading on aspects of literacy acquisition in a digloss situation. *Reading Research Quarterly*, 28, 70-79.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House/Harper Row.

Pienemann, M., M. Johnston, and G. Brindley. (1988). Constructing an acquisition-based procedure to assessing second language acquisition. *Studies in Second Language Acquisition*, 10, 217-243.

刘思，1995，英语听力词汇量与阅读词汇量——词汇研究调查报告，《外语教学与研究》第1期。

朱纯，1994，《外语教育心理学》，上海：上海外语教育出版社。

附录1：用于实验的10个生词

* rhythm 节奏，韵律； * crawl 自由式游泳；wrinkle 皱纹；display 陈列，展示； * stride 大步走，跨过； * spur 刺激，鼓舞；yacht 游艇，轻舟；dismay 沮丧，灰心； * stroke 一击，一划；combination 结合，联合

（注：有星号的为目标词）

附录2：用于词汇和听力理解测试的短文

Being able to swim has always been a valued skill. But competitive swimming is just over one hundred years old. Racing in the water began in Europe only about fifty years before the first modern Olympics. England was the leader in this new sport. Pools were built and races were held.

Interest was spurred when a group of native Americans went to England in 1844 to swim against an English team. To the surprise and dismay of the English, the Americans beat them easily. At the time the English used the breast stroke, which was thought a good form. The Americans used a kind of free style that was much faster.

A short time later an English family, who had moved to Australia, devel-

oped the crawl, an overhand swimming method. One of that same families also took it to America. Charles Daniels, an early user of the crawl, changed the kicking rhythm to develop the American crawl.

Listening comprehension questions:

1. When did racing in water begin in Europe?
   A. About fifty years before the first modern Olympics.
   B. Fifty years ago.
   C. When the crawl was developed.
   D. After the breast stroke was developed.
2. What did a group of Native American swimmers do in 1984?
   A. They learned the English way of swimming.
   B. They beat an English team easily in a competition.
   C. They introduced breast stroke to England.
   D. They failed in the competition.
3. Who developed the crawl?
   A. An Australian family
   B. An American family.
   C. An English family who moved to Australia.
   D. An Australian family who moved to England.

附录三：个案研究

(Note: The following is a multiple-case study in which six senior middle school students were asked to write a composition based on the pictures given while speaking out what was going on in their minds in order to find out how L1 was involved in the process of L2 writing and what are the specific functions.)

Thinking in L1 and L2 writing quality: A study of the process of English picture composition

Qiufang Wen, Nanjing University
Chunjie Guo, Nanjing University of Aeronautics and Astronautics

The paper reports on a study of the process of English picture composition in an attempt to reveal the relationship between thinking in L1 and L2 writing ability. The subjects involved in the study were 6 senior students from a secondary school in China. They were asked to write an English composition in about 100 words based on the pictures given, while speaking out whatever occurred in their minds by the method of think-aloud. The writing task was undertaken individually and the whole process of think-aloud was recorded. The data-analysis included two tasks: (1) rating the compositions; (2) transcribing the tapes and analyzing the protocols. Two native speakers were invited to score English compositions independently and the internal reliability between the two raters reached .83. The protocols were analyzed both quantitatively and qualitatively.

The results showed that the students with higher scores on composition used far less L1 than the students with lower scores in terms of the total amount. However, considering the five functions of L1 use identified from the data (i.e. transformation, confirmation, generating ideas, retrieving L2 forms and controlling the writing procedures), the students with higher scores employed more L1 for transformation and confirmation, but less L1 for the remaining three functions. Furthermore, the mean length of each L1 utterance used by the higher achievers was longer than that used by the lower achievers in the case of transfor-

mation and confirmation, but shorter than the lower achievers in the other cases. Due to the small sample size, the results yielded from this study will only serve as hypotheses to be tested in the future.

母语思维与外语写作能力的关系：对高中生英语看图作文过程的研究

南京大学 文秋芳
南京航空航天大学 郭纯洁

摘要：本文报告了笔者运用有声思维（think-aloud）的方法记录了6名高中生英文看图作文的过程，其目的是为了揭示母语思维与外语写作能力之间的关系。结果表明母语思维的参与量与写作成绩之间呈负相关。在英语写作过程中，母语的主要功能为转换中介、内容生成中介、形式检索中介、内容验证中介和程序管理中介。虽然高分组在总体上比低分组少用母语思维，但该组比低分组多使用中文作为转换中介和内容验证中介、少用中文作为内容生成中介、形式检索中介和程序管理中介。就使用中文的平均长度而言，高分组用于转换中介和内容验证中介的中文平均长度比低分组长，但用于内容生成中介和形式检索中介的中文平均长度比低分组短。由于样本人数少，以上结果只是为大样本的研究提供假设。

关键词：有声思维 母语功能 外语写作 二语习得

本文报告了笔者运用有声思维（think-aloud）的方法纪录了高中生英文看图作文的过程。该项研究的目的在于揭示母语思维与外语写作能力之间的关系。具体地说，该项研究首先研究作文得分与母语思维中使用中文词数的关系。第二步，考察母语在外语写作过程中的不同功能以及这些不同功能的动态交替。最后根据所建立的母语功能类别，比较高分组和低分组的差异。由于样本人数少，所得结果只是为今后的大样本研究提供假设。

一、研究背景

用外语写作文，特别在外语学习的初级阶段，以母语为中介似乎无法避免，然而大家又都认为，用母语思维会干扰外语学习的进步（John-

son，1985；Wen & Johnson，1997；Zamel，1982）。过去的研究着眼于作文中错误对比分析，以此推断母语的迁移或干扰作用。然而只研究输出结果，不探究输出过程，难以揭示母语参与外语写作的全部作用，更难界定母语参与思维在什么条件下产生促进作用，在什么条件下产生阻碍作用。

从80年代起对写作过程的研究日益增多（Arndt，1987；Brooks，1985；Diaz，1986；Jacobs，1982；Urzua，1987；Zamel，1983），其中大多数研究都试图检验第二语言的写作过程是否与第一语言的写作过程相同，只有少数研究涉及到母语对外语写作影响这一课题，其研究结果似乎都认为在外语写作过程中借助母语思维对提高外语作文的质量有积极的影响。例如，Lay（1982）拍摄了4个以中文为母语的外语学习者有声思维的写作过程，并就写作背景与态度与被试者进行了面谈。她研究的结果是：在写同一篇作文的过程中，运用母语思维多的学生比运用母语思维少的学生，在作文的内容、结构和细节上都表现出明显的优势。Zamel（1982）发现在她的8名研究对象中，惟一通过翻译来写作文的，英文写作水平最高，而其余的不仅不用翻译，而且认为翻译对英文写作没有帮助。Cumming（1987）在他的研究报告中指出：6个以法语为母语的加拿大成年人在用英文写作文时，写作水平低的只是通过法语生成英文作文的内容，而写作水平高的不仅用法语生成内容，检查文体，而且更多的是用于选择英文单词。根据他的看法，英文作文水平高的比水平低的对母语的依赖性更强。Friedlander（1990）研究了母语在外语作文内容的构思中所起的作用。他发现运用母语思维既不影响外语写作的时间，也不影响质量。Kobayashi & Rinnert（1992）比较了日本学生通过翻译写的英文作文和直接用英文写的作文，发现语言水平低的学生与高水平的学生相比，明显得益于翻译。

国内对外语写作过程中母语作用的研究几乎是一片空白，但最近郭纯洁、刘芳（1997）的“外语写作过程中母语影响的动态研究”颇具开拓性。郭和刘通过有声思维的方法研究了12名不同英语水平的学生看图作文的过程。根据对数据的定性分析，他们指出母语在外语输出过程中的主要作用有3个：1）对作文内容的逻辑推理；2）对语言形式的分析、判断；3）对相关外语词汇、短语或句子的检索。他们还指出逻辑推理的过程以母语思维为主。以上结果在过去的相关研究中均未涉及，这无疑对我们全面认识母语在外语写作过程中的作用有着重要的贡献，但该项研究只限于对母语动态影响的一般描述，对母语作用的界定过于

简单，不能用于解释所有使用中文的情况。另一方面，该研究还未涉及母语参与量的多少与写作成绩的关系。本文报告的研究就是对郭、刘收集的部分数据作更深入细致的分析，以比较高分组与低分组在外语写作过程中使用母语的差异。

二、研究设计

1. 研究对象

用于最后数据分析的6名学生来自安徽省临泉一中，其中高二和高三的学生各一半。他们选自20名被邀参加英文看图作文的高中生。挑选标准有两个：1）学生在面谈中对研究所用的“有声思维”方法态度积极，并认为这种方法基本上反映了他们写作时的思维活动；2）他们在有声思维的过程中沉默时间不超过总写作时间的20%。

2. 测量工具

用于英文看图作文的8幅连环画选自当地出版的教师参考书（见附录）。连环画描写的是一个学生在上学的路上如何帮助盲人过马路的故事。为了纪录学生有声思维的过程和所用的总时间以及沉默的时间，我们准备了一台高质量的录音机和计时器。

3. 数据收集

数据收集分二个阶段进行：1）“有声思维”方法训练；2）正式测试。为了使被试者明确有声思维的要求和获得有声思维的亲身体验，他们接受了必要的训练。首先由研究者进行有声思维示范，他一边读古文，一边说出自己理解古文的思维过程，并在示范基础上进行了必要的解释。接着，对学生试测。试测时，每个学生需用有声思维的方法完成两项任务：一是读一段古文，二是辨认一个不完整的图画或汉字。这两项试测任务经过精心挑选，均不涉及两种语言的交替使用，目的是希望被试者不要误认为在英文看图作文时必须要中英文交替使用。

正式测试的要求包括：1）根据所给的图画，用英文写出一篇100字左右的文章。2）必须自始自终说出写作过程中头脑中出现的任何想法。想法是以汉语出现，就说汉语；想法以英文出现，就说英文。说的过程中，不允许有较长的停顿和沉默。3）要边说边写。研究者对每个学生

看图作文时的有声思维情况不仅认真地做了观察纪录，而且进行了录音。

4. 数据分析

数据分析涉及两项任务：1）对作文进行评分；2）逐字纪录有声思维的情况，并对其进行分析。我们聘请了在南京大学任教的两位外籍教师，对被试者的6篇作文分别进行了等级评定。他们根据对文章的内容、结构和语言3方面的总体印象，将6篇作文的质量由高到低排出名次等级，6级为质量最高，1级为质量最低。两位教师所给成绩相关系数为.83（$p = .04$）。每个学生作文的最后等级根据二位教师所给等级平均数的排序。表1列出了每个学生的作文得分情况：

	教师A	教师B	平均等级	最后等级
学生1	2	1	1.5	1
学生2	5	5	5.0	5
学生3	3	2	2.5	3
学生4	6	6	6.0	6
学生5	1	3	2.0	2
学生6	4	4	4.0	4

表1：6名学生英文作文得分情况

对有声思维录音数据的分析分3步进行。首先，我们统计每个学生在看图作文的思维过程中使用中文词数占总词数的比例，然后计算该比例与写作成绩的关系。中文词的确定根据商务印书馆1996年出版的《现代汉语词典》修订本。统计中英文总词数时，没有排除重复出现的词。

第二步，我们从每个学生有声思维的纪录中分离出一系列使用中文的事件，每一个事件必须与某个功能相联系。每个中文事件功能的界定以郭和刘（1997）提出的3大功能为基本依据，但在此基础上作了较为详细的补充和归类，目的是为了能描述有声思维中出现的所有使用中文的事件。建立类别的过程比较费时耗力，原因是类别的建立需要反复修订和验证。第三步，根据建立的母语功能类别，比较高分组和低分组的差异，以此推断在外语写作过程中母语产生积极和消极作用的条件。

三、结果与讨论

该项研究的结果主要用于回答下列3个问题。第一个问题是母语参与思维量的多少与外语写作能力之间的关系。根据以前的研究结果，我们的假设是：作文得分高的学生比得分低的学生多依赖于母语思维；或者说多用母语思维的学生比少用的学生作文得高分的可能性更大。第二个问题是母语在外语作文的过程中究竟有哪些功能？这些不同的功能是如何在英语句子的写作过程中交替发挥作用的？第三个问题是作文得高分者和得低分者在使用母语上的差异如何体现在所建立的功能类别上。

1. 中文词的数量与作文成绩的关系

根据表2，高分组学生使用中文词数的平均比例（37%）大大低于低分组的学生（51%）。统计结果表明，中文词数所占总词数的比例与作文成绩之间的相关系数为 – .77（p = .07）。虽然相关系数比较高，但由于样本太小，不具有统计意义。这里负相关的含义是：写作思维过程中使用中文词数比例越高，学生作文得低分的可能性越大。由此我们可以提出假设：写作过程中多依赖母语为中介影响作文的质量；或者说英文作文水平低的学生不得不多用中文思维。究竟使用中文思维是导致作文水平高低的原因还是结果，相关分析不能回答这个问题。这里我们只能推断：写作思维过程中汉语使用比例的高低可以在很大程度上预测作文的得分。

上述结果刚好与 Lay（1982）、Zamel（1982）、Cummings（1987）和 Friedlander（1990）的相反。造成差异的原因可能是多方面的。其中重要的原因是，他们的结果都不是来自精确的量化数据。仅凭被试者自己的报告和研究者的印象，所得结论很可能不准确。当然这里不能排斥另一种可能性，即上述提到的研究，包括本文报告的研究在内，样本人数都很少，所得结果缺少可靠性。

		中文总词数	英文总词数	中文词数占中英文总词数的比例	作文等级排序
高分组	学生4	448	842	35%	6
	学生2	596	769	43%	5
	学生6	247	580	30%	4

（见下页）

（接上页）

		中文总词数	英文总词数	中文词数占中英文总词数的比例	作文等级排序
	平均数			37%	
低分组	学生 3	567	936	38%	3
	学生 5	587	524	53%	2
	学生 1	958	542	64%	1
	平均数			51%	

表 2：高分组和低分组使用中文词数比例的比较

2. 母语在外语看图作文过程中的功能

根据对有声思维数据的分析，母语在某一语言单位形成过程中，主要有 5 个功能：1）转换中介；2）内容生成中介；3）形式检索中介；4）内容验证中介和 5）程序管理中介（见表 3）。

母语功能名称	定义
转换中介	与后面紧跟着的英文有直接联系的中文，是作者转换英文的依据。
内容生成中介	在形成内容过程中所使用的母语，用于对内容的推理、判断、选择、监控、联想和提问。
形式检索中介	在形式检索过程中所使用的母语，用于对形式的判断、推理、监控、评价、联想和提问。
内容验证中介	把英文倒译成的母语，用于检查所用的英文形式是否表达了自己想要表达的内容。
程序管理中介	在决定写作程序时所用的母语。

表 3：母语功能的分类及其定义

下面我们首先从静态的角度，分别举例说明母语 5 种不同的功能；然后以动态的方式说明在外语句子/从句形成的过程中，这 5 种不同功能是如何交替发挥作用的。

1）对母语功能的静态描述

（1）转换中介

转换中介指与英文作文内容有直接联系的中文内容，是作者转换成外语语言形式的依据。判断转换中介的标准只有一个，即中文和与之相对应的英文内容紧紧连在一起，没有被其他与之没有联系的中英文所阻隔，例（a）和例（b）中划线的部分就是转换中介：

（a）吃过早饭以后，after having breakfast，after having his，after having his，after having his breakfast.（学生 4）

（b）上气不接下气 quite out of breath，可能上课过了吧 the class had begun.（学生 5）

（2）内容生成中介

所谓内容生成中介，就是指作者在表述作文内容生成过程中进行各种思维活动时所用的母语，例如用母语进行推理、判断、监控、联想、选择、提问等。它与转换中介的不同之处有两点。第一点是用于内容推理、监控、联想、选择和提问的母语，都与要写的外语作文没有直接联系，最终不进入转换阶段，但它有助于作者对图画的理解，最终形成要表达的内容；第二点是，用于判断内容的母语中虽有一部分与英文作文内容有直接联系，但与之相关的英文内容没有紧跟在相应的中文后面，而被其他中英文的思维活动所阻隔。下面用一些例子来说明这个问题。

在例（a）中划线部分是对内容的推理，它与英文作文内容没有直接的联系。斜体字部分“起来晚了”是对内容的判断，与之相联系的英文在后面出现了，但在“起来晚了”和“he woke up late”之间，有一系列与“起来晚了”没有直接联系的中英文交替活动。设想 he woke up late 紧接在“起来晚了”后面，“起来晚了”就是转换中介。例（b）和例（c）中的划线部分“还没有把他的意思表达出来”“这太罗嗦了”是作者对自己所写内容的评价或者监控；例（d）中的划线部分是作者对图画内容的联想，或者说是对别人行为的评价；例（a）中“这一幅画到底咋说呢?”和例（e）中“汤姆怎么跟他说话呢?”及“汤姆跟他说的是什么?”称为“内容提问”，这一类问题表明作者在内容的形成上碰到了暂时的困难；例（f）中画线的部分是对内容的选择。

（a）“肯定是来晚了，肯定是起来晚了（内容判断中介），要不然他咋能会，咋能会救了人，救了一个盲人就迟到呢（内容推理中介）。起来晚了，起来晚了（内容判断中介）。穿衣服，穿衣服

Tom put on clothes, clothes quickly... 从 clothes quickly、quickly 这再加上，加上从句吧 for he 用哪个系动词呢? for he, for he, wa wake up late, 算了, woke up, woke up late, woke up late.（学生 1）

(b) 还没有把他的意思表达出来（内容监控中介），这一幅画到底咋说呢?（内容提问中介）（学生 1）

(c) 到了学校之后，他发现已经晚了，and he found when go to his school, he run, 这太罗嗦了（内容监控中介）。（学生 2）

(d) ……车快来了，这开车的也是，你看不见前面有人吗? 哪为啥非要往前开呢? 道德品质极其恶劣，这个道德品质好（内容联想中介），小明过去了，冲上去 rushed up and pulled him back.（学生 5）

(e) 汤姆怎么跟他说话呢?（内容提问中介）汤姆要走了必定要跟他说句话（内容推理中介），汤姆跟他说的是什么?（内容提问）（学生 3）

(f) when he realized what had happened, he was, 当他意识到了发生了什么，可能很高兴（内容判断中介），还是又激动又高兴呢?（内容选择中介）（学生 5）

(3) **形式检索中介**

形式检索中介用于表述在外语形式检索过程中所产生的各种思维活动，其中包括对语言形式的判断、选择、监控、评价、提问和联想，下列例子中画线的部分说明了母语在此过程中的不同功能：

(a) ...a good deed 是复数还是单数?（形式选择中介）did a good deed、作了一件、一件好事（形式推理中介），对，did a good deed（学生 2）

(b) This time, as he was going to cross the street, He saw, He 应该小写（形式监控中介），这是一个主句，主句（形式推理中介）。（学生 4）

(c) 再加个，加个 with, 再加个 with（形式判断中介），no thought for his safety, with no thought for his own safety 怪好哩，正好管用，老师讲的这词组正好管用（形式评价中介）。（学生 1）

(d) 穿过人行道咋说呢?（形式提问中介）（学生 5）

(e) 用啥词儿呢? 他不顾自己的安危，对，上一次老师才讲的完形

填空。(形式联想中介)(学生 1)

(4) 内容验证中介

有声思维纪录揭示，作者有时在一个外语语言单位形成后，又把它倒译成中文，有时是全译，有时只是译其中的一部分。笔者设想，这里倒译的原因是为了要检验所形成的外语语言单位是否符合自己要表达的内容，因此这种倒译时使用的母语，我们就把它称之为内容验证中介，例如下列例子中画线的部分：

(a) almost strike him 差一点撞着他了(内容验证中介)，就在这一霎那间，冲上去，rushed up and pull him back 拉了回来(内容验证中介)。(学生 5)

(b) when he explained why he 他为什么来晚了 why he was late to his teacher 当他给老师解释他为什么来晚了(内容验证中介)，when he explained，... .(学生 3)

(c) When Tom passed through the street，he suddenly saw 突然看见(内容验证中介)，suddenly saw a blind man 一个盲人(内容验证中介)(学生 2)

当然这里不能排斥另一种可能性，即作者倒译是为了照顾前后内容的联系。由于当时没有在有声思维结束后立即和被试者面谈这个问题，因此这里的分析只能是一种假设，有待今后进一步研究。

(5) 程序管理中介

程序管理中介用于决定写作的程序，例如作者用母语思考究竟先写，还是先看下一幅图画；究竟先修改，还是先往下写等活动，此时使用的母语称之为程序管理中介，它与英文作文内容没有联系。由于被试者要求完成的是看图作文，就其程序上的思考在整个写作过程中并不多，但笔者考虑到对母语功能描述的完整性，还是将其单独列为一大类。另外出于将来研究的需要，学生用外语进行命题作文时，用于程序管理的母语肯定是不可缺少的一部分。请看下面例子中画线的部分就是我们所说的程序管理中介。

(a) 先别忙，先写上一点(程序管理中介)。想起来，想起来好像，电视里边，跟电视里边放的普通生活中，肯定是……(学生 1)

(b) 马上再修改，先写完(程序管理中介)a car went up with high speed(学生 5)

2）母语功能的动态描述

根据学生有声思维的数据，我们概括出母语在外语句子写作过程中的动态功能（见图1）。如图一所示，学生接受了图画刺激后，可从阿拉伯数字标出的三种选择中挑选任何一种：（1）图画——→外语；（2）图画——→转换中介；（3）图画——→内容生成中介，开始英语句子的写作。图中所有箭头都用虚线连接，以表示箭头所标出的路径只是可选择的路径之一。凡是箭头标向带虚点的方框，就表示写作过程中有母语参与，其具体功能如框中所注。箭头标向带实线的方框，表示语言形式是外语。双向箭头表示所标出的路径可以形成环路。由于篇幅的限制，我们不可能用例证说明图中标出的所有可能组合的路径。下面我们以例子说明三种途径产生英语句子/从句的过程中母语功能交替使用的情况。

在例（a）中，作者选择了路径1，即在3次接受了图画刺激后，头脑中首先出现的都是英语：（1）After he arrive，got to his school，when；（2）he found that；（3）class has，class，classes，class has，class had，class begin。有趣的是这三个英语语言单位的后面都出现了中文。“当他”和“他发现”是对相应英文的部分或全部倒译，目的在于验证所要表达的内容。思维中出现的“过去完成时”是对形式的监控。

（a）After he arrive，got to his school，when 当他（内容验证中介），when he got to school，he found that 他发现（内容验证中介），he found that，class has，class，classes，class has，class had，class hegin，he found that 过去完成时（形式监控中介），he found that class had，had hegun.（学生2）

例（b）中作者选择的是路径2，即看了图画后，将要表达的内容先用中文表述出来，然后再以中文为中介分两次转换成英文。当第二次转换结束以后，作者又把英文倒译成中文。

（b）他又穿过马路，就在他上学的路上，他要穿过一条马路（转换中介）on her，on his way to school，way，to school，s－c－h－o－o－l，on his way to school，he 他要穿过一条马路（转换中介）he had to，he had to cross，c－r－o－s－s，he had to cross a street 他要穿过一条马路（内容验证中介）（学生5）。

例（c）的作者选择了路径3：看到图画以后，没有用中文直接形成要表达的内容，而是用中文对图画的内容进行分析，推理以及对写作程序作出决定（见划线部分），其中除了“他往前走”以外，其他部分都

与英文作文的内容没有直接关系。但他将“他往前走”转换到一半时，对“when”的用法产生了疑问，于是用中文作为形式判断和推理的中介，解决了“when”的问题。此后作者又对内容“toward school”产生了问题，他先认为可以用 toward school，但立刻又否定了这一想法，提出的理由是“离学校还远着呢”，这时作者不能马上想出解决的方法，只好决定了“先这样写”。

(c) 他到底是在等汽车还是在穿马路，这有个路标，路标？哦，不对，这他的腿在往前走着哩。可能是往前走，可能是往前走，正当，可能是正当他，正当他，正当他，正当，他慢慢走，前面有 quickly，他现在慢慢走？（内容推理中介）一会儿再改，先写上，先写出来再讲（程序管理中介）。他往前走（转换中介），when he、he、用 when 好些（形式判断中介），when 表示突然（形式推理中介），he was walking，he was，he was walking，he was walking toward school、toward school，用 school 现在就说（内容判断中介），不对（内容监控中介），离 school 还远着呢（内容推理中介），先这样写（程序管理中介），toward school（学生 1）

从上述 3 个例子可以看出，无论学生选择哪一种途径，几乎都逃脱不了母语的参与。

3. 高分组和低分组在使用母语功能类别上的差异

在问题 1 中，我们已经得到结论：在英文看图作文的过程中，高分组学生使用母语参与思维的比例比低分组的学生低 14%，这一结果揭示了高分组和低分组使用母语思维的总体趋势，但我们并不能就此推断出这两组学生在使用每一类母语功能上的具体差异。因此，下面我们根据列出的母语功能类别，逐一比较这两组学生使用中文的情况，比较的内容包括 3 个方面：1）母语功能类别数量上的差异，即哪一个组使用的功能类别多；2）每个母语功能类别所用中文词数占中文总词数比例上的差异；3）母语功能出现的频率和使用中文的平均长度。

1）使用母语功能类别的数量差异

从母语功能的 5 个大类来看，高分组只有前 4 类，没有第 5 类（见表 4）。这就是说在看图作文的过程中，高分组中没有人用母语来描述写作的程序，而低分组用了 10 次，所用的中文词 46 个，占总中文词数的

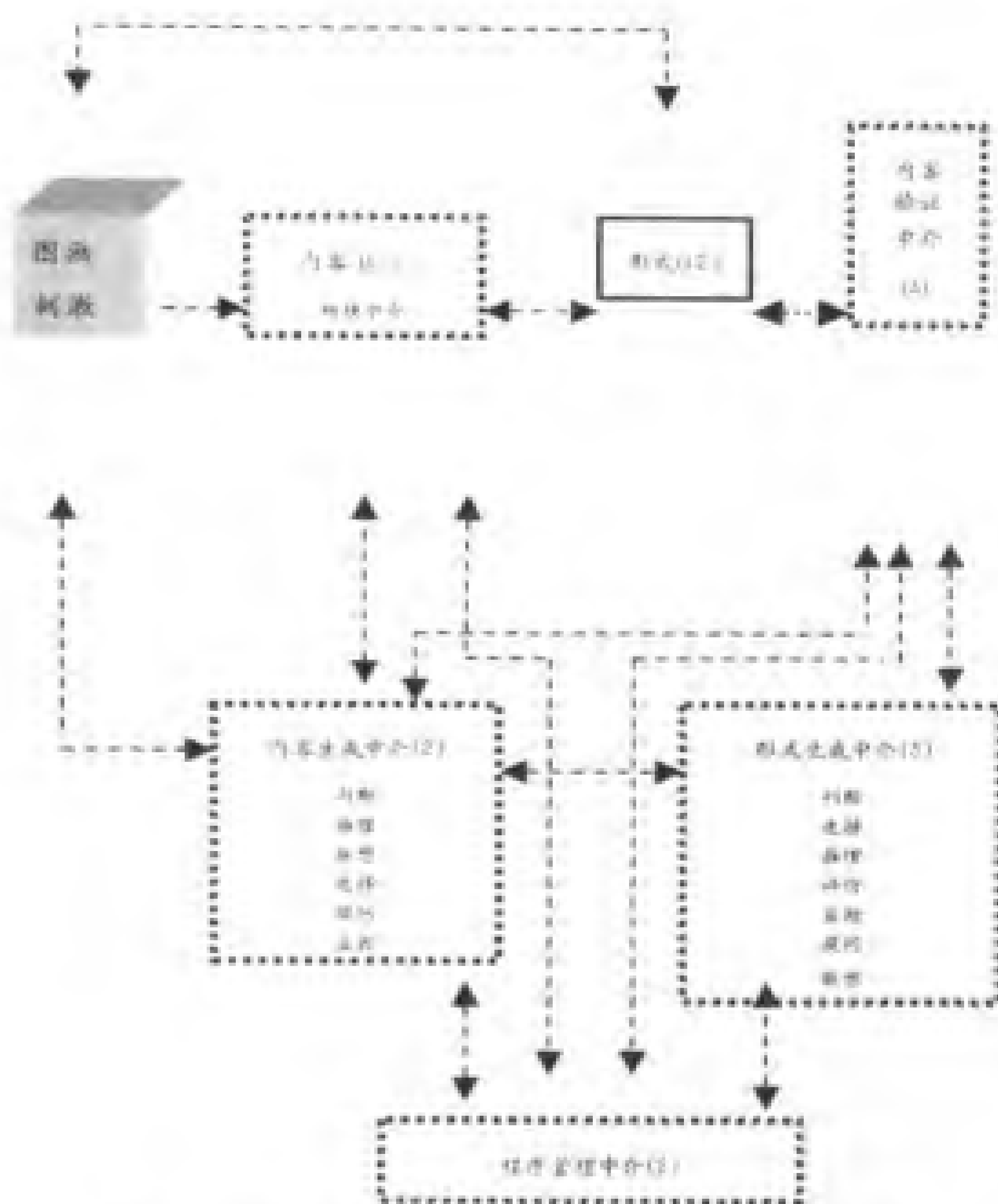


图 1：在外语句子写作过程中，母语的动态功能

2%。事实上，低分组需要对程序作出抉择，一般是因为对前面所写的内容发生了疑问，想修改但又不能立刻拿出很好的解决方法。

从内容生成中介和形式检索中介所包括的小类来看，高分组没有“内容选择”、“形式评价”和“形式联想”这 3 小类。问题是低分组在这几个小类上使用的中文词数占总中文词数的比例也很低。笔者猜想这很可能不是高分组和低分组之间的真正差异。如果样本扩大了，这种情况仍旧存在，我们才能比较有把握地作出结论：高分组使用母语的功能

类别比低分组少。

2）每个母语功能类别所用中文词数占中文总词数比例上的差异

表4中列出了某母语功能使用词数占中文总词数的百分比，具体的计算方法如下：

$$\frac{\text{高分组/低分组使用某个母语功能所用的中文词数}}{\text{该组使用中文的总词数}} \times 100 = \text{某个母语功能类别所占的百分比}$$

高分组在有的功能类别上所占比例比低分组高，在有的功能类别上，比低分组低。这一结果再不像问题一中所得到的结论那么简单了。高分组使用母语作为转换中介的占中文总词数的37%，而低分组只占15%；高分组在内容验证上使用的中文词数占中文总词数的16%，是低分组5%的3倍多；高分组用母语作为内容生成中介的比例只有28%，而低分组高达57%；用于形式检索中介和程序管理中介的比例，高分组略低于低分组。现在我们可以对问题一中作出的结论作进一步的修订，即在英文看图作文的思维过程中，虽然高分组总体上使用母语比低分组少，但在作为转换中介和内容验证中介上，比低分组使用的母语却多得多，少只是少用于内容生成中介、形式检索中介和程序管理中介上；虽然低分组在总体上使用母语比高分组多，但用于转换中介和内容验证中介上的母语比例却比高分组低得多。

在5个母语功能类别中，只有用于转换中介的母语跟英文作文的内容有直接联系，因此高分组多用母语作为转换中介，对英文作文质量的提高可能有积极的帮助。内容验证中介是用于检验所形成的外语形式是否与作者要表达的内容相一致，因此它可能对英文作文质量的提高也有促进作用。低分组在内容生成过程中使用了大量的母语，很显然这对写好英文作文帮助不大，因为所用母语中的绝大部分都与作文内容没有任何关系。

那为什么低分组要在这一类活动上花这么大气力和时间呢？是低分组不知道其中的奥秘吗？不能排斥有一部分学生确实缺少这方面的知识和技能，因此需要对他们进行必要的训练。笔者认为问题可能不会如此简单。低分组在内容生成过程中多用母语只是现象，其真正的原因可能是，低分组对图画的理解能力以及用母语写作的能力不如高分组，因此在内容的组织上需要用母语进行反复推敲。当然联想活动的产生，不能说明内容生成上有困难，但过多的联想至少说明作者不知道如何将有限的时间用在写作上，很可能他们在母语写作时也有类似的情况。上述原因仅是笔者的推测，需要进一步的研究。

3）使用中文的平均长度

表4分别列出了高分组和低分组使用每个母语功能的频率以及使用中文的平均长度。所谓使用中文的平均长度就是指每次使用某个母语功能时平均需要的中文词数，具体计算方法为：

$$\text{使用中文平均长度} = \frac{\text{使用某个功能使用的总中文词数}}{\text{某个功能出现的频率}}$$

母语功能类别		低分组				高分组			
		使用中文的总词数	某母语功能使用词数占中文总词数的百分比	某母语功能出现的频率	平均长度	使用中文的总词数	某母语功能使用词数占中文总词数的百分比	某母语功能出现的频率	平均长度
1. 转换中介		307	15%	79	3.9	463	37%	87	5.3
2. 内容验证中介		115	5%	35	3.3	208	16%	56	3.7
3. 内容生成中介	1）判断	433	20%	43	10.1	184	13%	27	6.8
	2）推理	323	15%	20	16.2	59	5%	6	9.8
	3）选择	26	1%	4	6.5	0	0	0	0
	4）联想	147	7%	11	13.4	11	1%	2	5.5
	5）监控	182	9%	32	5.7	85	7%	18	4.7
	6）提问	88	4%	18	4.9	32	2%	6	5.3
		1199	56%	128	9.4	371	28%	59	6.2
4. 形式检索中介	1）判断	182	9%	56	3.3	139	11%	50	2.8
	2）选择	48	2%	13	3.7	17	1%	6	2.8
	3）推理	40	2%	6	6.7	35	3%	8	4.4
	4）评价	45	2%	11	4.1	0	0	0	0
	5）联想	14	1%	2	7.0	0	0	0	0
	6）监控	54	3%	17	3.2	30	2%	12	2.5
	7）提问	62	3%	18	3.4	28	2%	7	4
		445	22%	123	3.6	249	19%	83	3.0
5. 程序管理中介		46	2%	10	4.6	0	0%	0	0
合计		2112		375	6.2	1291		285	4.8

表4：高分组与低分组在使用母语功能类别上的差异

以高分组的转换中介为例。根据表 4，高分组使用转换中介的频率为 87 次，共用中文词数 463，那么就转换中介而言，使用中文的平均长度为 5.3 个词，即 463 除以 87。就转换中介和内容验证中介来说，低分组的中文平均长度分别为 3.9 和 3.3，低于高分组的 5.3 和 3.7。在使用内容生成中介和形式检索中介时，低分组的中文平均长度分别为 9.4 和 3.6，高于高分组的 6.2 和 3.0。由于两组在内容验证中介和形式检索中介上使用的中文平均长度差异不大，因此不作进一步讨论。从逻辑上说，转换中介使用的中文平均长度越长，得到与之内容相近的英文形式也可能越长。如果上述推理能够成立，那么高分组转换的效率就比低分组高。为什么这两组的学生在这方面表现不一样？原因值得进一步研究。内容生成中介就其功能而言，是为了解决内容形成过程中碰到的问题，作者用中文进行推理、判断、选择、监控、联想和提问。所用中文的平均长度越长，说明问题解决得越不顺利，花的时间越多。从这个意义上说，低分组解决问题的能力比高分组低，其原因已在前面分析过，这里不再重复。

四、结论

该项研究的结果可以归纳如下：

1. 从总体上看，外语写作能力高的学生比能力低的学生对母语依赖性小，但高能力组比低能力组多使用中文作为转换中介和内容验证中介；而低能力组比高能力组多用中文作为内容生成中介、形式检索中介和程序管理中介。就使用中文的平均长度而言，高能力组用于转换中介和内容验证中介的中文平均长度比低能力组长，但低能力组用于内容生成中介和形式检验中介的中文平均长度比高能力组长。由此可见，写作能力高的学生对母语的使用是有选择的。
2. 母语在外语看图作文过程中一般作为转换中介、内容生成中介、形式检索中介、内容验证中介和程序管理中介，它们在外语句子/从句形成的过程中交替出现。
3. 学生接受了图画刺激后，可以有 3 种选择开始英语句子的写作：1）图画⇒外语；2）图画⇒转换中介；3）图画⇒内容生成中介。学生无论选择哪一种途径，几乎都避免不了使用母语。

以上结果，由于样本数量小，只能作为今后研究的假设。此外该课题还可以从以下几方面作进一步深入研究：1）将有声思维和面谈结合，让学生自己陈述使用母语的原因；2）研究学生进行英语命题作文时母

语参与思维的情况；3）将大学生英语看图作文的情况和高中生进行比较；4）对学生写作过程进行必要的指导是否能控制母语的消极作用，而发挥其积极作用。

参考文献

Arndt, V. (1987). Six writers in search of texts: A protocol-based study of L1 and L2 writing. ELT Journal, 41(4), 257-67.

Brooks, E. (1985). Case studies of the composing processes of five "unskilled" English-as-a-second-language writers. Dissertation Abstracts International, 47(1), 164A.

Cumming, A. (1987). Decision making and text representation in ESL writing performance. Paper presented at 21st Annual TESOL Convention, Miami, FL.

Diaz, D. (1986). The adult ESL writer: The process and the context. Paper presented at the 76th Annual NCYE Convention, San Antonio, TX.

Friedlander, A. (1990). Composing in English: effects of a first language on writing in English as a second Language. In B. Kroll (Ed.). Second Language Writing. Cambridge: Cambridge Applied Linguistics.

Jacobs, S. (1982). Composing and coherence: The writing of eleven pre-medical students. Linguistics and Literacy, 3. Washington, DC: Center for Applied Linguistics.

Johnson, C. (1985). The composing process of six ESL students. Dissertation Abstracts International, 46(5), 1216A.

Kobayashi, H. & C. Rinnert. (1992). Effects of first language on second language writing: Translation vs. direct composition. Language Learning, 42(2), 183-251

Lay, N. (1982). Composing processes of adult of ESL learners. TESOL Quarterly, 16(2), 406.

Urzua, C. (1987). "You stopped too soon": Second language children composing and revising. TESOL Quarterly, 21(2), 279-304.

Wen, Qiufang & R. K. Johnson. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. Applied Linguistics, 18(1). 27-48. Oxford University Press.

Zamel, V. (1982). Writing: The process of discovering meaning. TESOL Quar-

terly, 16(1), 195-209.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. TESOL Quarterly, 17(1), 165-187.

郭纯洁，刘芳（1997），外语写作过程中母语影响的动态研究，《现代外语》，第4期。

附录四：随机数字表

12651	61646	11769	75109	86996	97669	25757	32535	07122	76763
81769	74436	02630	72310	45049	18029	07469	42341	98173	79260
36737	98863	77240	76251	00654	64688	09343	70278	67331	98729
82861	54371	76610	94934	72748	44124	05610	53750	95938	01485
21325	15732	24127	37431	09723	63529	73977	95218	96074	42138
74146	47887	62463	23045	41490	07954	22597	60012	98866	90959
90759	64410	54179	66075	61051	75385	51378	08360	95946	95547
55683	98078	02238	91540	21219	17720	87817	41705	95985	12563
79686	17969	76061	83748	55920	83612	41540	86492	06447	60568
70333	00201	86201	69716	78185	62154	77930	67663	29529	75116
14402	53536	07779	04157	41172	36473	42123	43929	50533	33437
59911	08256	06596	48416	69770	68797	56080	14223	59199	30162
62368	62623	62742	14891	39247	52242	98832	69533	91174	57979
57529	97751	54976	48957	74599	08759	78494	52785	68526	64618
15469	90574	78033	66885	13936	42117	71831	22961	94225	31816
18625	23674	53850	32827	81647	80820	00420	63555	74489	80141
74626	68394	88562	70745	23701	45630	65891	58220	35442	60414
11119	16519	27384	90199	79210	76965	99546	30323	31664	22845
41101	17336	48951	53674	17880	45260	08575	49321	36191	17095
32123	91576	84221	78902	82010	30847	62329	63898	23268	74283
26091	68409	69704	82267	14751	13151	93115	01437	56945	89661
67680	79790	48462	59278	44185	29616	76531	19589	83139	28454
15184	19260	14073	07026	25264	08388	27182	22557	51501	67481
58010	45039	57181	10238	36874	28546	37444	80824	63981	39942
56425	50996	86245	32623	78858	08143	60377	42925	42815	11159
82630	80466	13592	60642	17904	99718	63432	88642	37858	25431
14927	40909	23900	48761	44860	92467	31742	87142	03607	32059
23740	22505	07489	85986	74420	21744	97711	36648	35620	97949

32990	97446	03711	63824	07953	85965	87089	11687	92414	67257
05310	24058	91946	78437	34365	82469	12430	84754	19354	72745
21839	39937	27534	88913	49055	19218	47712	67677	51889	70926
08833	42549	93981	94051	28382	83725	72643	64233	97252	17133
58336	11139	47479	00931	91560	95372	97642	33856	54825	55680
62032	91144	75478	47431	52726	30289	42411	91886	51818	78292
45171	30557	53116	04118	58301	24375	65609	85810	18620	49198
91611	62656	60128	35609	63698	78356	50582	22505	01692	36291
55472	63819	86314	49174	93582	73604	78614	78849	23096	72825
18573	09729	74091	53994	10970	86557	65661	41854	26037	53296
60866	02955	90288	82136	83644	94455	06560	78029	98768	71296
45043	55608	82767	60890	74646	79485	13619	98868	40857	19415
17831	09737	79473	75945	28394	79334	70577	38048	03607	06932
40137	03981	07585	18128	11178	32601	27994	05641	22600	86064
77776	31343	14576	97706	16039	47517	43300	59080	80392	63189
69605	44104	40103	95635	05635	81673	68657	09559	23510	95875
19916	52934	26499	09821	97331	80993	61299	36979	73599	35055
02606	58552	07678	56619	65325	30705	99582	53390	46357	13244
65183	73160	87131	35530	47946	09854	18080	02321	05809	04893
10740	98914	44916	11322	89717	88189	30143	52687	19420	60061
98462	89822	71691	51573	83666	61642	46683	33761	47542	23551
60139	25601	93663	25547	02654	94829	48672	28736	84994	13071

（Black. K. 1992：A2-A3）